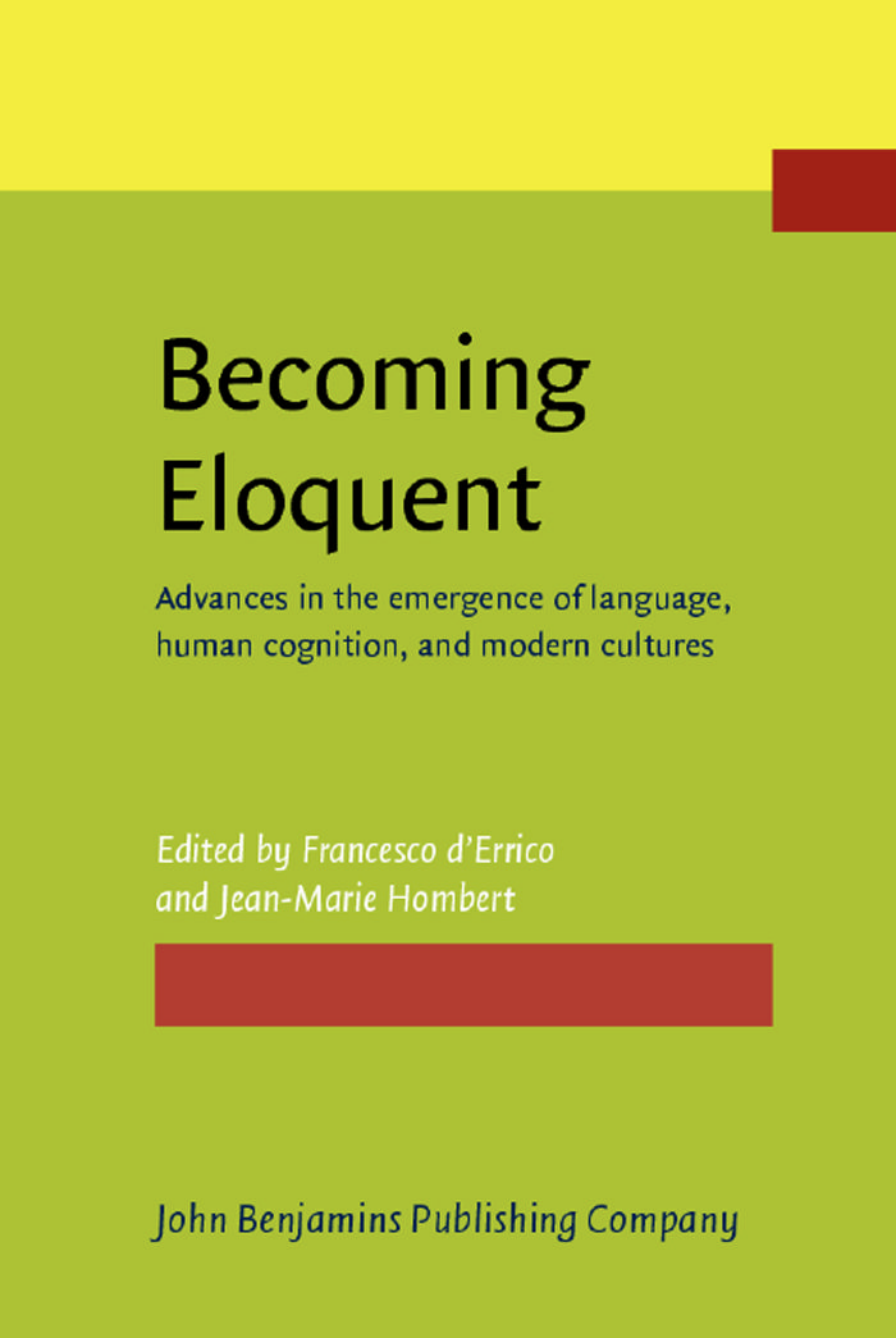




Becoming Eloquent

Advances in the emergence of language,
human cognition, and modern cultures

*Edited by Francesco d'Errico
and Jean-Marie Hombert*



John Benjamins Publishing Company

Becoming Eloquent

Becoming Eloquent

Advances in the emergence of language,
human cognition, and modern cultures

Edited by

Francesco d'Errico

CNRS & Université Bordeaux 1

Jean-Marie Hombert

CNRS & Institut des Sciences de l'Homme

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Becoming eloquent : advances in the emergence of language, human cognition, and modern cultures / edited by Francesco d'Errico, Jean-Marie Hombert.

p. cm.

Includes bibliographical references and index.

1. Anthropological linguistics. 2. Linguistic paleontology. 3. Language and languages--Origin. I. D'Errico, Francesco. II. Hombert, Jean Marie.

P35.B43 2009

306.44--dc22

2009036668

ISBN 978 90 272 3269 4 (HB; alk. paper)

ISBN 978 90 272 8871 4 (EB)

© 2009 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

Introduction	1
<i>Jean-Marie Hombert and Francesco d'Errico</i>	
Language and archaeology	
From the origin of language to the diversification of languages: What can archaeology and palaeoanthropology say?	13
<i>Francesco d'Errico, Marian Vanhaeren, Christopher Henshilwood, Graeme Lawson, Bruno Maureille, Dominique Gambier, Anne-Marie Tillier, Marie Soressi and Karen van Niekerk</i>	
Early diffusion of domestic bovids in Europe: An indicator for human contacts, exchanges and migrations?	69
<i>Anne Tresset, Ruth Bollongino, Ceiridwen J. Edwards, Sandrine Hughes and Jean-Denis Vigne</i>	
Language and genes	
Linguistic, cultural and genetic perspectives on human diversity in west-central Africa	93
<i>Lolke Van der Veen, Lluís Quintana-Murci and David Comas</i>	
The Berber and the Berbers: Genetic and linguistic diversities	123
<i>Jean-Michel Dugoujon, Clotilde Coudray, Antonio Torroni, Fulvio Cruciani, Rosaria Scozzari, Pedro Moral, Naima Louali and Maarten Kossmann</i>	
East meets West: The Assam corridor	147
<i>François Jacquesson</i>	
Genetic and linguistic diversity in Central Asia	163
<i>Evelyn Heyer and Philippe Menecier</i>	

Genetic and linguistic borders in the Himalayan Region	181
<i>Thirsa Kraaijenbrink, Emma J. Parkin, Denise R. Carvalho-Silva, George L. van Driem, Guido Barbujani, Chris Tyler-Smith, Mark A. Jobling and Peter de Knijff</i>	
Language acquisition and language universals	
From babbling to first words in four languages: Common trends across languages and individual differences	205
<i>Sophie Kern, Barbara Davis and Inge Zink</i>	
Language and animal communication	
The primate roots of human language: Primate vocal behaviour and cognition in the wild	235
<i>Klaus Zuberbühler, Karim Ouattara, Anderson Bitty, Alban Lemasson and Ronald Noë</i>	
Language evolution and computer modeling	
Can agent-based language evolution contribute to archeology?	267
<i>Luc Steels</i>	
Index	287

Introduction

Jean-Marie Hombert and Francesco d'Errico

Background

Few topics of scientific enquiry have attracted more attention in the last decade than the origin and evolution of mankind and language. Few have offered scholars an equivalent intellectual challenge and an equal opportunity for interdisciplinary collaboration from such a wide range of disciplines. This collective endeavour, however, has rarely been recognised as a necessarily joint venture by research funding agencies and encouraged by them through the launching of specific programmes. The EUROCORES programme “The Origin of Man, Language and Languages (OMLL)” of the European Science Foundation represents a remarkable exception to this rule.

The aim of the European Collaborative Research Scheme (EUROCORES) is to enable researchers in different European countries to collaborate in developing a scientific synergy in areas where European scale and scope are required to reach the critical mass necessary for top class science in a global context. EUROCORES provides a flexible framework that allows national basic research funding and participatory organisations to join forces to support quality European research in and across all scientific areas.

Traditionally, the study of the origin of language was considered too speculative and insufficiently anchored in empirically based studies to merit serious scientific attention. However, around the time of the launch of the OMLL programme, new data had been collected by scholars from several disciplines which had profound implications for the origin of this distinctly human means of communication. New perspectives opened by genetics, evolutionary anthropology, neurophysiology, and cognitive science were expected to converge and offer a solid ground for a fresh approach to the old problem of the origin of language(s).

Discoveries in archaeology and paleoanthropology, grounded in a more reliable chronological framework, thanks to advances in radiometric and luminescence dating methods, were providing pertinent evidence to test novel evolutionary scenarios. Following the pioneer intuition of L. Cavalli Sforza, comparative

maps of genetic and linguistic human families were produced, showing similarities between the distribution of genetic diversity and that of linguistic groups while, at the same time, challenging oversimplifications in this field of study. Similarly, the development of linguistic skills was seen to be linked to the evolution of the brain and of cognitive strategies.

In sum, towards the end of the 20th century, the study of the origin of language and of languages was emerging as a promising field for multidisciplinary research, where prehistoric archaeology, palaeoanthropology, genetics, linguistics, neurophysiology, cognitive science, as well as computer science and robotics, could profitably collaborate, and where international collaboration was expected to provide great benefits.

The launching of the OMLL programme in early 2001 marked the end of an intensive preparation phase and the beginning of a unique research scheme crossing boundaries not only between very diverse disciplines, but also within Europe. Sixteen national funding organisations from twelve different European countries brought together a programme budget of six million Euros to create a platform for pan-European research on the question of the origins of human language and the current distribution of languages and language families across the globe.

The programme invited proposals on six different themes: Language and Archaeology, Language and Genes, Language Acquisition and Language Universals, Language and Brain, Language and Animal Communication, Language Evolution and Computer Modelling. After an international peer-review process managed by the ESF, 21 Collaborative Research projects were launched in 2003. These 21 collaborative research teams consisted of 44 individual research projects based in 12 different European countries. Bringing together expertise from such a broad range of disciplines and such a wide variety of countries with different research traditions has been a unique opportunity for the participating researchers and students. Now that these networks have contributed concrete results advancing our knowledge of the emergence of human language, it is clear that these collaborations must be continued and strengthened in order to answer new challenging questions which have emerged from these recent studies.

The contributions presented in this volume are intended to depict a panorama of the topics explored by various research teams in the framework of the OMLL programme, to summarize new relevant data and emerging theories, to provide an updated view of this interdisciplinary venture, and, when possible, to provide directions for future research. In their variety, the collected papers cover the original call for proposals and reflect the multiplicity of interests and research strategy used to tackle this complex issue.

It is obvious that the papers included in this volume do not cover all aspects of research in the vast scientific area of the EUROCORES Programme "The Origin of

Man, Language and Languages". The editors are conscious that some areas have been inadequately covered and/or are not represented in this volume. This is especially true for instance in the case of Historical Linguistics. In this area, the European Science Foundation has recently funded a special workshop "New directions in Historical Linguistics" addressing specifically the relevance of recent research in Historical linguistics and the dispersal of languages. Contributions from this workshop are available at <http://25images.ish-lyon.cnrs.fr/Portails/NDHL/>.

Language and archaeology

Speech does not fossilize but the interpretation of artefacts obtained from archaeological sites can be used to infer the degree of complexity of the communication system necessary to produce these artefacts and sometimes their associated behaviour. The degree of complexity of tool technology is often used as a marker for the level of cognitive ability but it is difficult to extrapolate the need of a sophisticated system of communication from tool technology alone. One reason is that the ability to manufacture complex tools can be learned from observation and imitation without explicit tutoring requiring the use of language. Intentional burials and sea faring are activities that appear to require a greater reliance on linguistic communication for the purpose of enacting rituals and solving problems.

The proximity between Neanderthals and anatomically modern humans has been debated since the discovery of the first Neanderthal fossils. At first, the Neanderthals were classified as our direct ancestors with limited cognitive abilities. The current view is that the Neanderthals belong to a genetic branch which separated from the human lineage some 500,000 years ago. Recent studies have also shown that Neanderthals had greater cognitive abilities than previously thought (more complex tool technology, more sophisticated hunting techniques). Were these new cognitive accomplishments the results of contacts with groups of anatomically modern humans or were they Neanderthal innovations? The two theories are currently under discussion and evaluation. If Neanderthals were able to exhibit complex behaviours, what was the nature of their communication system? Contrary to what has been accepted since the 70's, their peripheral speech production system was probably not very different from ours.

In their paper **d'Errico, Vanharen, Henshilwood, Lawson, Maureille, Gambier, Tillier, Soressi and van Niekerk** review the evidence suggesting that the traits used to define behavioral modernity are not specific to our species. Instead the authors argue for a very gradual emergence of these traits over a long period of time on different continents and among different populations, including Neanderthals. The evidence evoked in their demonstration comes from the use of pigment,

engraved or painted representations, personal ornaments, burial practices, musical traditions, anatomy and encephalization. The authors' argument contradicts the hypothesis of a symbolic revolution coinciding with the arrival of anatomically modern humans in Europe some 40,000 years ago and highlights inconsistencies in the anatomically – culturally modern equation and the potential contribution of anatomically “pre-modern” human populations to the emergence of these abilities.

The so-called “Neolithic revolution” has had a major impact on population growth and human migrations. It involved domestication of plants and animals. In their article **Tresset, Bollongino, Edwards, Hughes and Vigne** examine the early diffusion of bovids in Europe as a trace of human migrations, contacts and exchanges. By investigating the localisation of wild ancestors and process of dissemination of different species they show that different processes took place. Sheep and goats had no wild ancestors in Europe, consequently domesticated sheep and goats can be considered as good tracers for the expansion of farming advance or at least of Neolithic influence. Cattle and pigs on the other hand had potential ancestors in Europe which means that independent local domestication or hybridization with animals imported from the Near East was possible. Genetic data indicate that domesticated sheep, goats and cattle were introduced in Europe with no interaction with local populations for goats and sheep or very little in the case of bovines. For pigs, the process was very different; a first introduction from the Near East was followed by a local domestication of the wild boar leading to the disappearance of the original Eastern lineages.

Language and genes

Genetic data from modern populations have been used to provide dates for the emergence of anatomically modern humans and to locate their continent of origin. They have also been used to infer information concerning more recent population movements. On the basis of linguistic data collected from modern languages it is possible to reconstruct earlier groupings of currently spoken languages into language families and thus infer population movements associated with these language groupings. In this volume, four geographical/linguistic areas are investigated: West-Central Africa, Northern Africa, Central Asia and the Himalayas. These studies illustrate the complexity of the relationship between genes and languages and clearly show that a better understanding of such historical scenarios will only be possible with the joint contribution of linguistic and genetic research.

Van der Veen, Quintana-Murci and Comas focus on West-Central Africa and address issues related to the so-called “Bantu expansion” and more specifically to

the peopling of the Cameroon-Gabon area. In this region, two populations with different lifestyles - farming villagers on the one hand and small groups of hunter-gatherers on the other - have been closely living together for a long period of time.

In their study, linguistically and culturally based hypotheses about specific relationships between Bantu-speaking groups and the peopling of the area over the last few millennia were compared to the results of genetic analyses (maternally-inherited mitochondrial DNA and paternally-inherited Y-chromosome variation). The analysis of mtDNA diversity did not show a clear correlation between languages and genetic markers for the Bantu-speaking agriculturalists. However, there are significant genetic differences between the agriculturalist ("Bantu") populations and the semi-nomadic hunter-gatherer populations (the so-called "Pygmies"). The mtDNA analysis strongly suggests (i) an initial divergence of the ancestors of contemporary "Pygmies" from an ancestral Central African population starting not earlier than ~70,000 years ago, (ii) a period of isolation between the two groups needed to explain their phenotypic differences, and (iii) long-standing and asymmetrical maternal gene flow from Pygmies to (proto) agriculturalists, starting not earlier than ~40,000 years ago and persisting until the last few thousands of years. The Y-chromosome study on the other hand suggests a Bantu to Pygmy flow of paternal lineages (i.e., introduction of male "bantu" villagers into the hunter-gatherers genetic pool). Jointly these results illustrate a sexually-asymmetrical mating pattern between "Pygmies" and "Bantu" and a strong process of language replacement in this region.

The objectives of the article presented by **Dugoujon, Coudray, Torroni, Cruciani, Scozzari, Moral, Louali and Kossman** are to correlate linguistic, historical and genetic studies from Berber languages and populations. In order to understand the peopling of North Africa, they compare the mtDNA diversity of Berber populations with geographically related populations (in particular from South-Western Europe, the Middle East and Sub-Saharan areas). The results show that Berber populations have, in the total mitochondrial diversity, an intermediate position between European and Sub-Saharan populations. More precisely, the data show a genetic differentiation between North-Western and North-Eastern Berber groups: populations from the Maghreb are related to European and Eastern populations (from Siwa in Egypt) and share more affinities with East and Sub-Saharan populations. The authors argue that one of the elements necessary for understanding the possible links between genetics and linguistics is the linguistic analysis of the time depth and geographical spread of a language. Some of the problems encountered in this respect in the specific case of Berber in Northern Africa are discussed, and some approximate indications on the chronology and geography of the Berber homeland are given.

The three following papers examine the relationship between the people and languages around the Himalayan range.

Jacquesson describes the ethnic and linguistic situations in the linguistically complex Assam corridor (North-East India). He stresses the importance of geographical locations, specifically the role of altitude, access to food resources and population densities. Looking at the process of language maintenance and language replacement, Jacquesson highlights the difficulty of the interaction between linguistic, political and economic constraints in this region.

Heyer and Mennecier's contribution shows hows population genetics facilitates the reconstruction of the history of population in Central Asia, especially Uzbekistan, Tajikistan, Kyrgyzstan. A comparison between linguistic and genetic data is also provided. Seven ethnic groups (Tajiks, Yagnob, Kyrgyz, Karakalpaks, Uzbeks, Kazaks and Turkmen) were sampled. The results of this study show that the genetic diversity clusters in two groups: Indo-European populations and Turk-Mongol populations (except for Uzbek). There is a good correlation between linguistic and genetic data especially with the Y-chromosome analysis (paternal line).

Kraayenbrink, Parkin, Carvalho-Silva, Van Driem, Barbujani, Tyler-Smith, Jobling and de Knijff examine the linguistic vs. genetic situation in the Nepal-Bhutan area. Three major questions are addressed: (1) Is there a correlation between language, genes, and geography in the Himalayan region?; (2) Is it possible to determine the genetic relationships (ancient ancestors) of the Bhutanese and deduce possible migration routes?; and (3) What can be said about the relative ages of the various groups if one compares the few real «aboriginal» groups with the others?

The authors collected genetic data from about 2000 individuals in Nepal and Bhutan and compared their results with those obtained on individuals from China and India in order to establish connections with outside genetic/linguistic groupings. They also find a good correlation between linguistic groupings and genetic similarities with a boundary running east to west, south of the border between India and Bhutan and through Southern Nepal. This border separates languages (and populations) speaking Indo-European languages vs. Tibeto-Burman speaking populations. Given the available genetic evidence, it is difficult to argue that any Nepalese or Bhutanese language group has been the genetic source for other Asian populations. Thus, it seems highly unlikely that there has been a major "out-of-Bhutan" or "out-of-Nepal" migration to either the south (India) or the north (Tibet/China). They conclude that the populations located today in Nepal and Bhutan probably originated outside their current locations but further genetic analyses will be necessary to propose more precise homelands.

Language acquisition and language universals

It is widely acknowledged that the «ontogeny-recapitulates-phylogeny» position is an oversimplification of evolutionary processes. In the realm of language origin, researchers have been reluctant with few exceptions to extrapolate results from language acquisition in children into possible evolutionary scenarios of language origin. This does not mean, however that ontogenetic studies are irrelevant to the understanding of evolutionary processes but it does mean that, when conducting research in this field, one ought to keep in mind the different contexts of language acquisition and language origin – language developing in a linguistically rich environment and a rapidly developing brain as opposed to a poor or non-existent linguistic context and a possibly a slowly evolving brain.

The emergence of complexity in human language can be compared with potential correlates of emerging complexity in contemporary ontogeny. The aim of the contribution of **Kern, Davis and Zink** is to compare the developmental trajectory of speech production capacities in children acquiring different languages from the babbling period to the emergence of early grammar. This paper focuses on the question of phonetic continuity between babbling and the early lexicon by asking the following questions: Do all children follow the universal trends described in the literature? Are sound patterns in babbling the same as those used in first words across language environments? The consistent data collection procedures adopted in the paper are designed to distinguish common trends vs. individual and language differences. The languages investigated are Dutch, French, Romanian, and Tunisian Arabic. Lexical and phonetic characteristics are longitudinally and crosslinguistically analyzed from various point of view: cumulative vocabulary, consonant place and manner of articulation, vowel types, and intrasyllabic and intersyllabic co-occurrence patterns. Despite prominent individual differences, the majority of children tended to follow the common trends reported in the literature. To address the question of the influence of language input, the authors compared child language frequencies to patterns in a 1,000 word dictionary data set for each language. Results indicate that input does not play an obvious role on observed patterns across the chosen languages for either consonants or vowels. In several other areas there is some evidence that perceptual learning from input influence the children's production-output patterns. In these cases, prevalent trends in the adult form of the languages studied are also those preferred by the majority of the children. However, it should be noted that in these cases the children's preferred trends are also favoured throughout the languages of the world. The data provide an answer to the question of continuity across babbling, first word, and later word periods. Results indicate that phonetic complexity expands after the lexical spurt due in part to a diversification and increasing complexity in

speech characteristics. This increase in complexity could be related to increasing control over speech articulators, which enable increase in capacities for matching language forms

Language and animal communication

Language is often used as a central distinctive feature of our species. Consequently, it is crucial to be able to provide a detailed account for the emergence of this specific behaviour. Is it just a system comparable but different from other animal communication systems or does it constitute a radically different way of communicating with other members of one's own species? A fast growing literature on animal communication clearly indicates that animals are capable of communicating information and exhibiting behaviour far more complex than previously thought.

Zuberbühler, Ouattara, Bitty, Lemasson and Noë examine the communicative abilities of several species of monkeys and apes and draw a comparison with the development of anatomical, neural, behavioral and cognitive features of human language. Wild groups of monkeys and apes were observed during their daily activities in order to understand what communication signals they are able to produce, under what circumstances they produce them, and what sorts of responses they elicit from listeners. Subsequently recorded samples of particular calls were presented to naïve receivers. For example, chimpanzees produce different types of food grunts depending on the type of food they find, e.g. differentiating between apples or bread. These calls are shown to be meaningful to other chimpanzees in the sense that if the receiver listens to "apple grunts" he is more likely to look for food in places where he previously found apples, but not bread, and vice versa. In conclusion, this comparative study shows that a number of properties of human language are also found in the communication systems of non-human primates. Individuals of these species are able to encode information about an external event and can also modify the meaning of messages as a function of their audience.

Language evolution and computer modelling

Computer modelling has been used quite extensively in recent years to simulate the evolution of the human vocal tract and the emergence of sound systems, lexicon, and syntax.

The notion of self-organisation in complex systems was first applied to the field of biology. Recently, it has been found to be relevant to the field of linguistics

where some encouraging results have been obtained for predicting the emergence of sound systems and linguistic structures in the evolution of communicative behaviours of our hominid ancestors.

In his paper, **Steels** presents the relevance of agent-based evolution models. His research shows how grammatical languages about complex scenes can emerge. After presenting the different types of models that have been developed, Steels focuses on socio-cognitive models. These models are necessarily complex if they aim at approaching reality; they have to take into account perception, memory, conceptualization, planning, joint attention, etc. The basic constituents of these models are the agents (i.e., a population of individuals), a world in which these agents operate and a type of interaction between these agents called a language game. In the author's simulations, agents have the capability to visually perceive dynamic scenes and analyze them in terms of event structures. Then they play a language game in which one agent describes to another a recently perceived event. In order to succeed in the game agents are required to invent and negotiate lexicon and grammar. Steels shows that an abrupt genetic modification is not a necessary step for the emergence of language but rather insists on the importance of the right social context and cognitive mechanisms.

Conclusion

In conclusion, the contributions collected in this volume do not claim to tell the last word on what was the first word. They fully demonstrate, however, that the origin and diversification of language has become a mature field of interdisciplinary study and that researchers working in this field are able to tackle general as well as technical issues. Funding of the EUROCORES OMLL programme was the necessary first step in developing fruitful long term collaborations across disciplines on these complex and multi-faceted issues related to the origin of language. New avenues of research have emerged through these collaborations and the scientific interest of a fast growing number of young researchers observed in recent years is clearly a very encouraging sign for the future.

Finally we wish to thank the European Science Foundation for its organizational and financial support and particularly Eva Hoogland for her efficiency and determination in putting together these contributions.¹

1. This work, as part of the European Science Foundation EUROCORES Programme OMLL, was supported by funds from the EC Sixth Framework Programme under Contract no. ERAS-CT-2003-980409

Language and archaeology

From the origin of language to the diversification of languages

What can archaeology and palaeoanthropology say?

Francesco d'Errico¹⁻², Marian Vanhaeren³,
Christopher Henshilwood²⁻⁴, Graeme Lawson⁵,
Bruno Maureille⁶, Dominique Gambier⁶, Anne-Marie Tillier⁶,
Marie Soressi⁷ and Karen van Niekerk⁸

In this paper we recall the arguments put forward in an attempt to link language origins and specific elements of the fossil record (pigment use, burial practices, personal ornaments, production of depictions and carvings, musical traditions, various anatomical features), and summarise the scenarios proposed by palaeoanthropologists and archaeologists to account for the emergence of modern behavioral traits. This review challenges the idea of a strict link between biological and behavioural change and suggests that modern cognition and language are results of a gradual, complex and non-linear process to whose advancement different human populations and possibly a number of fossil human species have contributed.

-
1. Université de Bordeaux, CNRS UMR 5199, Institut de Préhistoire et de Géologie du Quaternaire, France.
 2. Institute for Human Evolution, University of the Witwatersrand, Johannesburg, South Africa.
 3. Ethnologie préhistorique, Nanterre, France.
 4. Institute for Archaeology, History, Culture and Religion, University of Bergen, Norway.
 5. McDonald Institute for Archaeological Research, University of Cambridge, Cambridge, United Kingdom.
 6. Université de Bordeaux, CNRS UMR 5199, Laboratoire d'Anthropologie des Populations du Passé, Talence, France.
 7. Institut National de Recherches Archéologiques Préventives.
 8. Department of Archaeology, University of Cape Town, Rondebosch, South Africa.

Introduction

It is reasonable to think that in human evolutionary history there must once have existed societies with no language or with communication systems less complex than the one which represents the quintessential feature of our present-day species. After the studied neglect which for more than a century hampered the study of language origin, a number of theories have come to flourish in this field (see for ex. Lieberman 1984, Ruhlen 1994, Dunbar 1996, Jackendoff 1999, de Wall 2007). In spite of such valuable contributions, however, none of these theories has brought to the fore unequivocal evidence as to how and when this crucial human ability may have emerged and what evolutionary steps, if any, it might have followed to reach its present state of development. We argue here that one reason for this failure may be that not enough attention has been paid to the archaeological and fossil record.

In the absence of writing, language does not fossilize of its own accord. However, fossil evidence of such features as encephalisation and the evolution of the vocal tract may assist in determining the course and progress of the evolution of syntactic speech (Arensburg *et al.* 1990, Martinez *et al.* 2004). Complex technologies, regional trends in the styles and decoration of tools, systematic use of pigments, abstract and representational depictions in a variety of media, mortuary practices, grave goods, musical instruments, and personal ornaments: these are amongst the more common, long-lasting human creations that may be considered, to varying degrees, as non-linguistic phenotypes associated with the emergence of language (Marshack 1976, Aiello 1990, Donald 1991, Klein 1999, 2000, Klein & Edgar 2003, Barham 2002, Mellars & Gibson 1996, Noble & Davidson 1996, d'Errico *et al.* 2003, Henshilwood & Marean 2003, Henshilwood *et al.* 2004, Conard 2005, Vanhaeren & d'Errico 2006, Chase 2006, d'Errico & Vanhaeren 2009, Henshilwood & Dubreuil 2009, Számadó *et al.* in press). They could, if their significance in this respect was more precisely evaluated, provide valuable information as to the origin and evolutionary steps that have led to modern language.

In this paper we will summarise the scenarios proposed by palaeoanthropologists and archaeologists to account for the origin of modernity; we will suggest how these disciplines can test these scenarios and will recall the arguments put forward, in an attempt to link language origins and specific elements of the fossil record. We will then highlight a number of relevant facts, many of them borrowed from the research we have conducted in Africa, Europe, and the Near East, within the framework of the programme *The Origin of Man, Language and Languages*. We will show how some of our results support an early presence of non-linguistic phenotypes for languages in Africa and the Near East. We will show how other results challenge the idea of a strict link between biological and behavioural change suggesting that modern cognition and language are results of a gradual, complex and non-linear

process to whose advancement different human populations and possibly a number of fossil human species have contributed. Finally we will point out some methodological deficiencies in previous studies and suggest some research perspectives.

How to test scenarios for the origin of modernity and language

The emergence of symbolism and language has for a long time been considered by archaeologists and palaeoanthropologists, as well as a number of linguists, as the result of a sudden change, taking place in Europe 40,000 years ago and coinciding with the arrival in this region of *H. sapiens* (Mellars & Stringer 1989, Stringer & Gamble 1993, Mellars 1996a, b, Mithen 1996, Bar-Yosef 1998, 2002, Conard & Bolus 2003, see Klein 1999, 2000 for a slightly different scenario). This model, known as the Human Revolution scenario, has been gradually replaced in the last decade by a new paradigm, called the Out-of-Africa scenario (McBrearty & Brooks 2000). According to this scenario, the process that produced our species in Africa (Cavalli-Sforza *et al.* 1994, Barbujani 2003, Templeton 2005, Ingman *et al.* 2000, Forster 2004, Li *et al.* 2008) must have granted it a number of advantages – syntactical language, advanced cognition, symbolic thinking – that favoured its spread throughout the world, determined its eventual evolutionary success and the extinction of pre-modern human populations with little or no biological and, if any, little and unbalanced cultural interaction. Underlying the Out of Africa model for the origin of modern behaviour is the view that the emergence of each of these new features marked a definite and settled threshold in the history of mankind and that the accumulation of resulting cultural innovations (more advanced technology, better adaptation, symbolic activities) contributed, as with genetic mutations, to create human societies increasingly different from those of their non-modern contemporary counterparts. If this model is to be correct such cultural innovations should not to be found outside the African continent before the arrival of modern colonisers. Postulating that these advantages were determined by a biological change leads logically, however, to the somewhat paradoxical conclusion that archaeology does not inform us as to the origin of modern behaviour and language. Populations will be considered smart, eloquent and symbolic according to their taxonomic status and not on the basis of the material culture they have left behind. It has been argued (d'Errico 2003, Jehs 2004, d'Errico & Vanhaeren 2007, 2009, Villa *et al.* 2005) that to avoid this pitfall archaeologists should adopt a large scale comparative approach. Documenting and dating the occurrence of these innovations in various regions of the world including Eurasia, the alleged realm of pre-modern populations, may reveal their presence at times and places incompatible with the Out-of-Africa model. It may also show a discontinuous pattern with

innovations appearing and disappearing or being associated in a way that does not match the trend predicted by the original “Out-of-Africa” scenario. This process can eventually lead to a revision of a simplistic “Out-of-Africa” scenario and the proposition of a model for the origin of modern cognition that best fits the empirical evidence or, at least, explicitly outlines facts in contradiction with this scenario, and that need explanation.

A similar hypothesis-testing approach should in our view also be applied to the human fossil records. Ancient human remains are rare and often of ambiguous taxonomic attribution (see for ex. Wolpoff 1999, Trinkaus & Zilhão 2002, Trinkaus *et al.* 2003, Henry-Gambier *et al.* 2004, Trinkaus 2005, Tattersal & Schwartz 1999, Tillier 2005). It can therefore be misleading to attribute or deny language abilities to past humans just on the basis of their taxonomic affinities. In order to avoid the danger of circular reasoning, palaeoanthropologists should develop methods to evaluate language, and more in general cognitive abilities, by relying on anatomical characters directly involved in those performances rather than on hypotheses shaped by close, but different, disciplines such as genetics or evolutionary psychology.

Language and the archaeological record

The arguments which have so far been put forward to link cultural innovations to the emergence of language vary both in nature and in interpretive value. The most common argument visualises human experience through the lens of uniformitarianism, the principle which represents the foundation of the Natural Sciences. Most archaeologists assume that cross-cultural analysis of historically known human societies can identify regularities that may help shed light on the way people thought, communicated, and acted in the past (Binford 1983, Gardin 1990, Renfrew & Zubrov 1994, Renfrew 1996, Tschauer 1996, Roux 2007). Prehistoric societies which display in their social and cultural systems the same degree of complexity recorded in historically known societies must have had means of communications which are comparable in complexity (d'Errico *et al.* 2003, d'Errico & Vanhaeren 2009, Henshilwood & Dubreuil 2009, Számadó *et al.* 2009, but see Botha 2008).

Knapping and bone working techniques which require complex sequences of actions have been considered viable correlates for the presence of behavioural modernity and complex language on the grounds that those techniques would have required a structured “syntax” of actions to be produced (Dibble 1989, Deacon 1997, Stout *et al.* 2000, McBrearty & Brooks 2000, Henshilwood *et al.* 2001b; Roux & Bril 2005, Villa *et al.* 2009, but see d'Errico *et al.* 2003, Backwell & d'Errico 2005, Stout & Chamina 2007).

Connections between music and language origins have traditionally been established on the grounds that the two systems share, to some degree, comparable syntax and semantics (Lerdahl & Jackendoff 1983, Thomas 1995). Such a parallel has been developed and refined more recently by Brown (2000) and Patel (2008). According to Brown, music and language could have evolved as two specializations from a common ancestor, called ‘musilanguage’. The model predicts that the rise of music as we know it (combining syntax based on pitch blending and pitch patterning to generate complex sound-emotion relationships) would have been accompanied by the development of a language with symbolic capacity for representation and communication. Drawing on similarities observed in the domains of rhythm, melody, syntax, meaning and evolution, Patel proposes convincingly that clear computational links exist between music and language.

Both theories imply that if we were able to identify any complex, socially shared and transmitted musical traditions in the archaeological record, these might be seen as a proxy for the presence of a complex language. This, however, is a highly contentious prediction.

In contrast to these approaches, behavioural psychologists such as Pinker (1997, 2000) think that our species’ remarkable musical capacities are fortuitous and largely inconsequential by-products of language-driven aptitudes; that “music is ‘auditory cheesecake’: a behaviour which parasitizes various other domains of human experience and no more than a technology that our ancestors developed simply because it gave them pleasure”. Yet no such scenario has been proven within the historical music-language relationship, and its development through archaeological time is still untested. How science chooses to define ‘music’ remains largely a question of perspectives and preferences which are constructs of our own time. We are thus free, if we wish, to join with Cross (1999) in arguing quite the opposite of Pinker’s case: suggesting ‘language’, and not ‘music’, to be the dependent development.

Part of the problem, of course, is that ‘music’ is notoriously difficult to define; or rather, that there are as many definitions of music as there are scholarly viewpoints and opinions. However, for our present purposes a reasonable, broad definition of ‘music’ – as distinct from ‘language’ – might be *the human-initiated generation and non-syntactical manipulation of non-lexical sounds (of whatever kind and by whatever means) using manufactured tools, naturally available objects or the human body – including especially the human voice*. Even here an obvious caveat must be that such ‘non-lexical, non-syntactical’ qualification cannot entirely exclude lexical and syntactical elements – they are, for example, perfectly proper constituents of text-based song. Music can also *include* other forms of symbolic meaning: by imitation of natural sounds, for example, or through the establishment of referential connexions between particular sequences or textures of sounds

and particular moods, behaviours, events or things; but it does not need to do so to be 'music'.

There are remarkable *commonalities* between 'musical' and linguistic behaviours today: indeed in some respects the two are sufficiently interwoven, not to say compounded, for investigators of modern populations to be able to sustain several different, highly plausible yet totally conflicting, ontological models. Can these commonalities afford us reliable musical proxies for symbolic language as expressed in speech and signing? There are several reasons why they might. Firstly, 'music' and 'dance' are clearly expressions of language, of a kind, insofar as they are *forms of expression in which sentiments or moods may be expressed through motorically differentiated and temporally organised gesture* (in sound and/or movement). They may also incorporate symbolism. Even though, in the 'pure' abstract forms in which taxonomists like to think of them, they are *non-verbal*, and therefore *not* obviously capable of conveying intelligent thought and argument (at any rate in as concise a manner as the symbolic languages of speech or signing), they utilise many cognitive and motor capacities in common with 'language'. Expression of each involves selection of, and thus discrimination between, different kinds of actions or gestures, and each involves synthetical, sequential organisation and development through time. Their discriminatory mode involves the selection and development of *individual* actions – in music's case *sounds*: sounds of different frequencies, amplitudes, durations and temporal profiles, to generate and modify different effects. Their synthetical organisation (in music's case, again of sound; in dance, of movement), is a process which follows *rules*, some universal, others culturally variable, as frameworks within which individuals may express themselves to each other in complex and extended ways. They can express single moods, whether briefly or continuously; or they can express different moods, and thereby, through sequential juxtaposition, *changes* of mood. With the addition of mimicry (which is a proto- or at least quasi-symbolic behaviour practiced by many modern animal species, not just living humans) they can readily take on the character of narrative.

Symbolic manifestations are the more cited evidence for the emergence of language and numerous scholars consider that the origin of language is linked intimately with that of the symbolic thought. A key characteristic of all symbols is that their meaning is assigned by arbitrary, socially constructed conventions (Chase & Dibble, 1987, Chase 2006). Perhaps the greatest benefit of symbolically mediated behaviour is that it permits the storage and display of information external to the human brain (Donald 1991, d'Errico *et al.* 2003, Wadley 2001, 2003, Henshilwood & Marean 2003, Hovers *et al.* 2003). It is widely accepted that a direct connection must exist between, on the one hand, the eminently symbolic character of human language today – with its capacity to refer to facts that are real and things that are imaginary, to the past and to the future – and, on the other, the creation,

maintenance and change of the material expression of symbolic thought in human cultures (Donald 1991, Aiello 1998, Mellars & Gibson 1996, Stringer & Gamble 1993, Mellars 1996a, b, 1998, Deacon 1997, Mellars *et al.* 2007). Only a communication system like human language or something functionally equivalent to it can unambiguously transmit the symbolic meaning of signs as well as the structured links between them. Also, because syntactical language is the only natural system of communication which bears a built-in meta-language, one that permits the creation and transmission of other symbolic codes (Donald 1991, Aiello 1998, Deacon 1997, Bickerton 2003), systematic use of symbolic material culture, when discerned in ancient human populations, may be seen as a reliable proxy for the acquisition of language and fully modern cognitive abilities.

Symbolic meanings can be attributed to elements of the natural world (e.g. humans, animals, landscape features) but such practices leave no detectable archaeological signature. When symbolic codes are embodied in material culture, the link between meaning and the referent becomes not only arbitrary but also, as with sounds in language, artificially created. This makes it possible for members of a society to locate symbolic items in strategic locations and spatially, if not syntactically, organise the links between these artificial symbols (d'Errico & Vanhaeren 2009). Apes are able to learn referential symbols and represent other minds in a context of social competition (Byrne 1995, Rumbaugh & Washburn 2003, Tomasello *et al.* 2005). Chimpanzees clearly have the capacity to develop and transmit cultural traditions (Whiten 2005), but they have never been observed creating systems of symbols in the wild nor embodying them in their material culture, or displaying them on their body.

Specific frames of inferences have been proposed to refine the link between language and particular elements of past symbolic material culture. A case in point is colour recognition and pigment use. Berlin and Kay's (1969) proposition that eleven basic colour words are common to all the languages of the world and that they arose in different languages in a regular sequence (black and white followed by red, green, yellow, blue and brown) has been used to support the view that colours did emerge in the same order during the process that has led to modern cognition (Hovers *et al.* 2003, but see Watts, in press). Since syntactical languages with only three basic colours exist, the identification of past societies specifically spotting red and/or black pigments, and purposely heating minerals to get a particular shade may be seen as an indication that not only did they have complex system of communication but also that those universals and semantic categories were already present in their languages.

The female cosmetic coalition model (Knight *et al.* 1995, 2008, Power & Aiello 1997; Watts 2009) posits that the use of red pigments for painting bodies and clothing, to mimic menstrual blood, should be seen as one of the strategies

created by females to build a more co-operative society in order to meet higher reproductive costs. Language is perceived, in this context, as the natural outcome of a new social framework dominated by the «rule of law». Although not making robust predictions about when complex languages would emerge along this process, this model clearly identifies a systematic use of red pigment as a proxy for syntactical language or, at least, for the social context in which such a mean of communication could emerge.

Some authors (Wadley 2001, 2003) counter a symbolic interpretation of early pigment by stressing the ambiguous character of this element of the material culture. The first often mentioned reason is that pigments is also used by human cultures for utilitarian purposes, which makes reaching a univocal interpretation of archaeological pigment a difficult task. Exclusively functional uses of pigment however are very rare (Watts 1999, 2009, in press) and may not exist at all. One could argue it may have been different in the past and that the recent widespread use of pigments for symbolic reasons in ethnographically attested traditional societies does not imply similar use by Neanderthals or other hominids. However, since this may be said of any other category of past material culture such a stand comes down to denying the possibility of inferring the past from the archaeological record.

A second more relevant ambiguity that characterizes pigments is that unlike depictional engravings, mobiliary art, personal ornaments, decoration on objects and cave walls, utilised pigments do not represent the direct outcome of past symbolic behaviours. A piece of ochre or manganese may have been ground, scraped or notched to produce the powder used to paint the Chauvet Cave, to decorate clothes, objects, or make body decorations on living or dead members of a group. It may also have been used directly as a crayon on these different surfaces. If the product of such actions don't survive in the archaeological record, utilised pieces of colorants will represent the only tenuous clues archaeologists can use to get an idea of those lost complex behaviours. And probably only some of them will survive. The amount of pigment recovered in an excavation depends on the excavation procedures used and taphonomic factors, including the technique used to prepare the pigment, the media on which pigment was applied, and the frequency of the activities in which pigments were used. The mere presence of utilized colorants may be an indicator that other pigments were also used but may not have survived. It is thus clearly crucial to document and examine all archaeological aspects linked to the use of pigments before automatically equating their discovery to symbolism.

A link between language and personal ornaments has also been proposed. It has been argued that symbolic items with no utilitarian purpose, created for visual display on the body, and the meaning of which is permanently shared by the members of a community represent a quintessential archaeological proxy for the use of language or, at least, of an equally complex communication system (d'Errico *et*

al. 2003, Vanhaeren 2005, Vanhaeren & d'Errico 2006, Henshilwood *et al.* 2004, d'Errico & Vanhaeren 2009, Dubreuil in press, but see Botha 2008). Symbols applied to the physical body ascribe arbitrary social status to the wearers that can be understood by the other members of the group only if the latter share the complex codes that establish a link between the worn items, the place and way they are displayed on the body, the social categorisation they signal and the symbolic meaning carried by the objects. No "institutionalised" symbolic meaning can be transmitted without language abilities (Searle 1995).

The earliest use of pigment

Over the past decade, systematic pigment use has been repeatedly mentioned as one of the defining elements of 'modern behaviour' (Knight *et al.* 1995, Watts 1999, 2002, in press a, b, Barham 1998, 2002, McBrearty & Brooks 2000, d'Errico *et al.* 2003, 2008, Henshilwood *et al.* 2002; Henshilwood & Marean 2003, Hovers *et al.* 2003; Henshilwood *et al.* 2009) and a proxy for language (Knight 2008, Watts in press a). Many instances of pigment use are reported from Early and Middle Pleistocene sites from Africa and Europe (e.g. de Lumley 1969, Leakey 1971, Marshack 1981, Wreschner 1981, McBrearty 2001). The supporting evidence, however, has rarely been illustrated and/or analysed in detail in publications. Exceptions are the Acheulean site of Bečov in the Czech Republic, and the Middle Stone Age (MSA) sites Twin Rivers in Zambia, Sai Island in Sudan (Van Peer *et al.* 2004), and Pinnacle Point in South Africa (Marean *et al.* 2007, Watts in press). At Becov four grinding stones used to rub and mix pigments and numerous fragments of porcellanite of various shades (yellow, orange, red), some bearing traces of use and heat-treatment, were recovered in an early Mousterian layer attributed to MIS 7 (200–250 ka) (Marshack 1981, Šajnerová-Dušková *et al.* in press) (Figure 1). The results of a replication experiment using grinding tools, similar to those found at Bečov, to process porcellanite, suggests that the archaeological porcellanite was purposely ground to obtain various shades of pigment powder.

Fieldwork at Twin Rivers in 1999 (Barham 1998, 2002), led to the discovery of 176 possible fragments of pigments in layers attributed to 260 ka and 400 ka. Five different pigment colours are recorded but only a few red pigments bear clear traces of use. Excavations at Sai Island (Van Peer *et al.* 2003, 2004) have yielded 59 yellow and red pigment lumps associated with worked sandstones and other grinding tools showing traces of pigments (Figure 1). The associated lithic assemblage is attributed to the Lower and the Middle Sangoan and is dated to ca 200 ka.

Robust evidence of an early systematic use of pigments comes from cave PP13B located at Pinnacle Point near Mossel Bay (South Africa). At this site 57



Figure 1. a: fragments of porcellanite found in the Early Mousterian levels of Bečov, Czech Republic; b: experimental production of pigment by grinding using a variety of raw material found at Bečov; c: grinding stone from the lower Sangoan levels of Sai Island, Northern Sudan; d: lumps of yellow pigment from the same levels. Scale bars = 1 cm (a-b after Šajnerová-Dušková *et al.* in press, c-d modified after Van Peer *et al.* 2003, 2004)

pigment pieces, weighting a total of 93.4 gr were recently found in layers dated to ca 164 ka (Marean *et al.* 2007). Ten pieces bear probable traces of scraping or grinding and two are interpreted as having been used but without proposing a more precise functional interpretation (Figure 2). Watts (in press) reports 380 pigments from this site, most of which are red or dark red in shade, found in layers dated to between 164 ka and 90 ka.

More than 8000 fragments of ochre among which ca 600 with clear anthropogenic modifications (Watts 2009, Henshilwood *et al.* in press) come from the MSA layers at Blombos Cave dated at 75 ka (Figure 2). Utilised ochre is particularly abundant in the M3 phase with an age of c. 100 ka (Jacobs *et al.* 2007).

After Marine Isotope Stage 6 ochre becomes a common feature at all MSA sites (Figure 2). The use of red pigments increases during the MSA and becomes a constant feature of MSA 2b, Still Bay, Howiesons Poort and MSA 3 sites (Wurz 2000, Watts 1999, Rigaud *et al.* 2006). According to Watts (1999, 2002, in press b) MSA people had a preference for the use of a strong red colorant even when yellowish or yellowish-brown material of similar chemical composition was available. These deliberate choices seem to contradict a purely functional interpretation for pigment use. Many MSA colorants are shaped as crayons, suggesting they may have been used to trace lines on soft material such as leather or for body painting.

Evidence indicating that the spread of pigment use is not limited to sub-Saharan regions comes from the site of Sodmein in Egypt (Van Peer & Vermeersch 2000). A levallois flake from levels dated to c. 115 ka is marked on both sides by a continuous red ochred line perpendicular to the flake main axis (Figure 2). In this case, however, the ochre staining may represent residues left by hafting the tool. More than one hundred fragments of red pigments (Figure 2), some bearing traces of use and heating, come from the Mousterian levels of Qafzeh site, Israel, dated to c. 100 ka (Hovers *et al.* 2003, Godfrey-Smith & Ilani 2004). This site has yielded burials attributed to *H. sapiens*, but no pigments have been found in clear association with the skeletons.

In Europe, more than 40 Mousterian sites dated to MIS 6–3 yielded objects described as blocks of pigments or stones used to grind or crush pigment (e.g. Bordes 1952, 1972, Demars 1992). The large majority of pigment used by Neanderthals is manganese dioxide, which produces a black pigment. Red pigment is rare. Most of these sites date from the end of the middle Palaeolithic, between 60 ka and 40 ka, and are attributed to the Mousterian of Acheulean tradition or the Charentian Mousterian. Ongoing analyses of materials by Soressi and d'Errico (2007) from two neighbouring sites, Pech de l'Azé I and Pech de l'Azé IV, the former, excavated in the 1960's by François Bordes, has delivered the largest known collection of Mousterian pigments (Soressi *et al.* 2007). More than 500 blocks of pigment come from this site (Figure 2 and 3), some found during the new excavations in a level older than 43 ka (Soressi *et al.* 2007). Most of Pech de l'Azé IV deposit was excavated during the nineteenth century, so it is likely that the 500 blocks and fragments being studied represent less than half of manganese fragments abandoned by Neanderthals. The upper layers of the nearby site Pech de l'Azé IV are contemporary with this Pech de l'Azé I (McPherron *et al.* 2001), but have yielded only 26 pieces within 9 archaeological levels: 15 of these show traces of modification. A few pieces of red and yellow ochre were also found at both sites, but none of them bear clear traces of use.

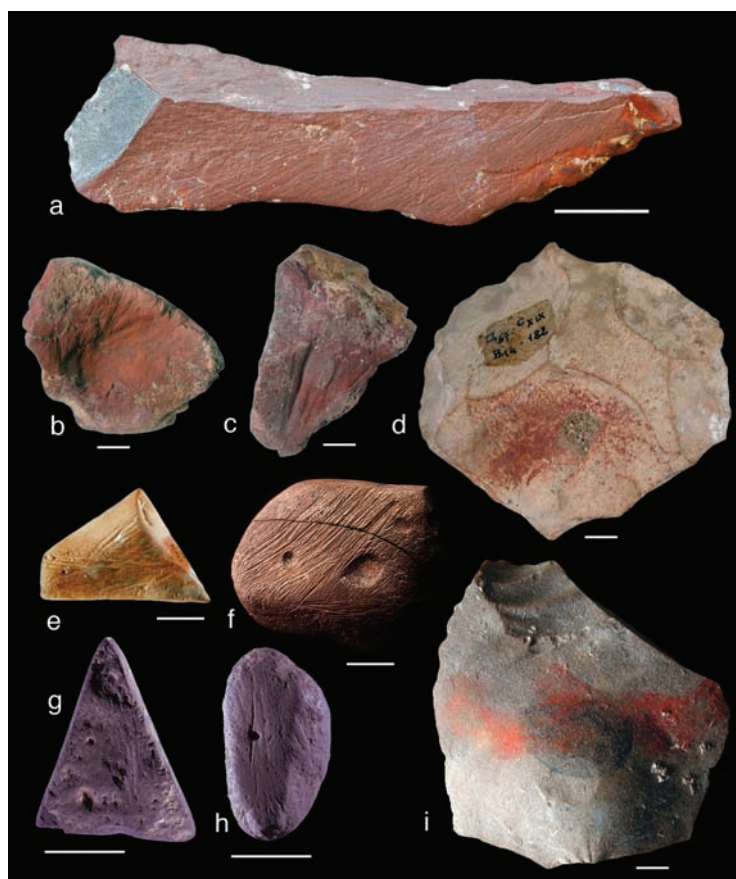


Figure 2. Modified pigments and ochred stone artefacts from cave PP13B at Pinnacle Point, Mossel Bay (a), the Mousterian levels of Qafzeh, Israel (b-d), the Middle Stone Age levels of Klasies River (e), Blombos Cave, South Africa (f), the Mousterian levels of Pech-de-l'Azé I France (g-h), and the Middle Stone Age levels of Sodmein cave, Egypt (i), (after Marean *et al.* 2007, Hovers *et al.* 2003, Wurz 2000, Henshilwood *et al.* 2001, Van Peer & Vermeersch 2000, d'Errico 2008). Scale bars = 1 cm.

Although manganese is available in the environment close to the site, the variety of shapes (slabs of various thicknesses, pebbles) and their composition suggests that Neanderthals were aware of all potential sources of raw material.

Analysis of the blocks reveals a clear difference between the natural surface of the piece, which is irregular, and areas used by humans, which are flattened by abrasion and appear polished in some cases. Few of these pieces appear to have been scraped or engraved by a sharp object like a flint edge or a piece of bone: 250 pieces from Pech de l'Azé I bear ground facets. A number of features show that

these modifications are due to the use of pigments and are not natural or post depositional in origin: a worn grinding stone stained with pigment come from the level richest in manganese fragments, the abraded facets are similar to those observed in MSA, Upper Palaeolithic and experimental ground pigments, and the location and morphology of wear surfaces are not random.

The narrow edges of the blocks were systematically used in order to produce elongated facets. To produce these facets, Neanderthals exerted a back and forth movement on a grinding stone, in a direction parallel or oblique to the axis of the future facet. This process has helped to produce flat or slightly convex facets. Seventy percent of the facets carry parallel striations visible to the naked eye. In the remainder the prominent areas between striations are smoothed and in some cases the surface of the facet has lost all trace of the original grinding.

Quantitative analysis of the microtopography of the archaeological facets and surfaces of manganese fragments experimentally modified or used in thirteen different ways suggests that after being ground Neanderthals used the manganese pieces to mark soft materials such as animal or human skin (Figure 3). The elongated shape of the facets on the archaeological specimens is consistent, as confirmed experimentally, with the intention of producing clearly visible straight black lines, perhaps arranged to produce abstract designs. The use of black pigment for body painting is consistent with recent palaeogenetic analyses of Neanderthal remains from El Sidron Cave, near Oviedo, Spain. Lalueza-Fox *et al.* (2007) found a variant of the Melanocortin 1R allele (MC1R) in two Neanderthals that is not present in modern humans. Its effect is to produce red hair similar to that seen in modern redheads. According to these authors Neanderthals probably had a light skin colour and the whole range of hair colour observed now in European populations. Further, recent genetic evidence (Krause *et al.* 2007) indicates that a critical gene known to underlie speech – namely FOXP2 – was present in the Neanderthal genome and that its appearance predates the common ancestor (dated to around 300–400 ka) of modern humans and the Neanderthals. This evidence helps support the contention that body decoration can be regarded as a proxy for language abilities.

The use of ochre becomes widespread in Europe after 36 ka during the Aurignacian, the technocomplex that is widely accepted as representing the first *H. sapiens* in Europe. Ochre is also found at sites attributed to other Early Upper Palaeolithic technocomplexes such as the Chatelperronian in France and the Uluzzian in Italy.

Substantial quantities of red and black pigments are for example found in the Chatelperronian layers of the Grotte du Renne, Arcy-sur-Cure, some of them clearly shaped by grinding to create crayons (Leroi-Gourhan & Leroi-Gourhan 1965, Salomon pers. comm.). The Uluzzian levels of the Cavallo cave, Apulia, Italy, have also yielded fragments of ochre and limonite (Palma di Cesnola 1993). Ochre



Figure 3. Top: manganese lumps from Pech-de-l'Azé I Mousterian site bearing flat facets produced by grinding and incisions made by stone tools; bottom: experimental use of ground manganese to colour human skin. Scale bars = 1 cm (photos F. d'Errico and M. Soressi)

crayons carved by scraping and decorated with sets of notches have recently been found at the Early Upper Palaeolithic site of Pekary Ila, Poland, in pre-Aurignacian layers (d'Errico & Vanhaeren 2007).

The human remains associated with the Chatelperronian indicate that this technocomplex is associated with Neanderthals (Bailey & Hublin 2006). This implies that at the end of their evolutionary history some Neanderthal communities were using both black and red pigments.

Finally red pigments appear in Australia since the very beginning of the human colonisation of this continent, probably between 50 ka and 60 ka, as evidenced by used pieces of ochre in the lowest levels of Australia's oldest sites such as Malakunanja II and Nauwalabila, Northern Territory (Taçon 2004). Ochre is also associated with one of the oldest Australian burials, Mungo III, found at Lake Mungo site and dated to ca 40 ka (Bowler *et al.* 2003).

Abstract and depictional engravings

Depictional and abstract representations are generally considered among the best examples of symbolic material culture. At present the oldest securely dated, purposely made engraving come from Blombos Cave. At this South African site two ochre slabs engraved with geometric patterns (Figure 4) come from MSA layers (M1 Phase) dated by two different dating methods to 75 ka years (Henshilwood *et al.* 2002). These two objects, the first published of a collection of fourteen, were found associated with Still Bay bifacial points, shaped bone awls and bone spear points (also in M2 phase dated at c. 78 ka) (Henshilwood *et al.* 2002). Possible engravings on bone (d'Errico *et al.* 2001, d'Errico & Henshilwood 2007) consisting of parallel striations and sets of joining lines come from M1 and M2 Phases. Both the published ochre engravings are variants of the same pattern suggesting they are not accidental "non-purposeful" behaviour. No functional interpretation of these engravings can be reasonably implied, as has previously been suggested for utilized ochre from Middle Stone Age sites (Klein 1999, Mithen 1999) and for pigments from Lower and Middle Palaeolithic sites (Mellars 1996, Wadley 2003). The Blombos engravings were each made on the edge of an ochre slab that was carefully prepared by grinding before use and in each case a lithic point was used throughout the engraving process.

A further twelve engraved ochre pieces from Blombos Cave, recently described by Henshilwood *et al.* (2009) comes from the M1, M2 and M3 phases, the latter with an age of ca. 100 ka (Jacobs *et al.* 2007). These engravings consist of parallel or joining lines (Figure 5) and criss-cross patterns of various types produced by single and multiple stroke motions, often on surfaces previously flattened by grinding. As

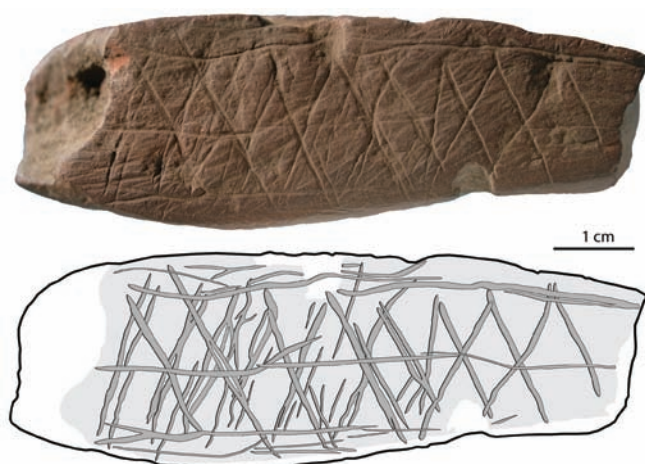


Figure 4. Photo and tracing of a thick ochre slab engraved with an abstract pattern from the Still Bay layers of Blombos cave dated to ca 75 ka years ago (photos and drawing F. d'Errico, after Henshilwood *et al.* 2009)

these engravings occur throughout the Blombos sequence, this suggests that symbolic intent was present in southern Africa at an earlier date than previously thought. A single engraved pebble with a cross-hatched design is recently reported as coming from the HP levels at the Klein Kliphuis site in the Western Cape, South Africa (Mackay & Welz 2008).

Engravings or notched pieces on bone and ostrich egg shells are reported from MSA contexts at other Sub-Saharan sites such as Klasies, Apollo 11, Hollow Rock Shelter, Border Cave, and Diepkloof (McBrearty & Brooks 2000, Rigaud *et al.* 2006). Our analysis of similar bone engravings from Klasies (d'Errico & Henshilwood 2007) suggests that in some cases notched bones best fit the interpretation of tools used in scraping activities, but in other cases a non-functional explanation seems more probable.

A number of abstract engravings on bone and manganese are reported from Acheulean and Mousterian sites in Europe and the Near East. Many have been shown to be natural. This is the case with the Pech-de-l'Aze II rib and several purported engraved bones from Cueva Morin, Stranska Skala, Schülen, Molodova IV, and other sites (d'Errico 1991, d'Errico & Villa 1997, Nowell & d'Errico 2007). Deliberate engravings, however, are apparently present at European Lower and Middle Palaeolithic sites. Cases in point are the mammoth shaft fragment from Bilzingsleben with a fan-like engraved motif (Behm-Blancke 1983, Steguweit 1999); the Tata "engraved" nummulite and polished mammoth dental plate (Marshack 1976), the parallel lines on the Temnata slab (Crémades *et al.* 1995); parallel



Figure 5. Left: photo and tracing of a thin ochre slab engraved with a set of joining lines from the Middle Stone Age layers of Blombos cave dated to 100,000 years ago. Right: close-up view of the engraved pattern. Scale bars = 1 cm. (photos and drawing F. d'Errico; after Henshilwood *et al.* in press)

incisions on bone or antler from Ermitage, La Ferrassie, Vergisson IV, Vaufrey; still unpublished shaft fragments with dozen of parallel lines from the late Mousterian levels of the French sites of Unikote, La Chapelle-aux-Saints, Marillac, Les Pradelles as well as a criss-cross patterns on a fragment from Peyrère (d'Errico *et al.* 2003, Soressi & d'Errico 2007, Vandermeersch *et al.* 2008, d'Errico unpublished data).

Potentially symbolic objects are sparse in Near Eastern Mousterian sites, despite the presence of *H. sapiens* in this region from c. 100 ka. The only two known examples are a cortex engraved with a set of concentric lines found at Quneitra (Marshack 1996, d'Errico *et al.* 2003) in a level dated to ca 60 ka and another cortex from Qafzeh with a set of parallel incisions found in the same levels as the burials (Hovers *et al.* 1997). Evidence comparable with the most ancient African symbolic expression is rare and the Berekhat Ram figurine is the only reliable example. It derives from an Acheulean layer sandwiched between two well-dated volcanic deposits, indicating that the human presence at the site probably lies between 250 ka and 280 ka (Goren-Inbar 1986). A microscopic analysis of the object combined with replicative experiments have shown that the object was purposely modified by humans (d'Errico & Nowell 2000) but this alone does not demonstrate its

symbolic nature. More questionable is the so called Tan-Tan “figurine” from Southern Morocco (Bednarik 2003), which has an uncertain stratigraphic provenance. In peri-Arctic regions there is an increase in the number of sites dated to c. 35 ka – 40 ka years, which yield Middle Palaeolithic or “transitional” stone tool industries associated with sequentially notched bone, perforated shells, and ivory working (Pavlov *et al.* 2001, Anikovitch *et al.* 2007). There is no evidence however that these assemblages were produced by *H. sapiens* and not by Neanderthals.

In Africa the oldest evidence for iconographic representations are the painted slabs with animal figures from the Apollo 11 site, Namibia, found in a layer dated to 26 ka – 28 ka (McBrearty & Brooks 2000). In Europe, the oldest painted and engraved representations are found at the Chauvet Cave in Ardeche; they are dated to 32 ka BP (Clottes 2003).

A buried fragment of rock painting, probably fallen from the ceiling of the shelter, has been found at Carpenter's Gap, North of Western Australia, in a layer dated to 39 ka (O'Connor 1995). Although undecipherable, because of its fragmentary nature, this painting confirms that pigment was used in Australia in symbolic activities and was applied on a variety of media since the continent was first colonised.

In summary, prior to 40 ka there are isolated depictions associated with Neanderthals in Europe, Neanderthals and modern humans in the Near East, and modern humans or archaic *Homo sapiens* in Africa. Iconographic depictions only appear only after 40 ka in Africa and Eurasia.

Personal ornaments

Ivory beads and perforated teeth and shells associated with the Aurignacian from Europe and the Near East, dated between 37 ka – 30 ka, were until recently unanimously considered the most ancient instances of personal decoration, but new discoveries, many of which were made by the members of our *Origin of Man Language and Languages* team, have challenged this perception in the last few years. We now know that marine shells were used as beads in the Near East, North Africa, and Sub-Saharan Africa by at least 82 ka. Six sites, Skhul and Qafzeh in Israel, Oued Djebbana in Algeria, Grotte des Pigeons in Morocco, Blombos Cave and Sibudu in South Africa have yielded evidence for the ancient use of personal ornaments.

The oldest securely dated personal ornaments come from the Grotte des Pigeons site near Taforalt, Morocco. Excavations conducted since 2003 by Barton and Bouzougar identified a 2.5 m thick stratigraphic sequence with five principal units (A, B+C, D, E and F), each of which is bracketed by a significant shift in sediment type (Bouzougar *et al.* 2007). Middle Palaeolithic occupation horizons have been recorded in units C-F. Unit E is characterized by Middle Palaeolithic

tools such as side scrapers and small radial Levallois cores, and a few thin, bifacially worked foliate points. Thirteen *Nassarius gibbosulus* shells (Figure 6) have been recovered from this unit, the majority (n=11) coming from contiguous squares covering an area of 6m². Seven shells come from a lightly-cemented 12 cm thick ashy lens with abundant evidence of human presence including archaeological finds and hearth debris. The presence of two beads in the overlying unit is attributed, in the light of the site formation process, to reworking due to human activity. The four remaining shells were found in the fill of burrows which intersects the ashy lenses, the layers from which they probably derive.

Grotte des Pigeons shells show features such bryozoa, tiny shells and sea worn gravel embedded into the body whorl, characteristic of dead shells accumulated on a shore. This demonstrates that the shells are not leftovers of human food. Possible evidence for the stringing of the perforated shells as beads comes from the identification on ten specimens from Grotte des Pigeons of a wear pattern different from that observed on modern reference collections and unperforated specimens from this site. Microscopic residues of red pigment, identified as iron oxides with a very high proportion of iron were detected on ten shells from this site. A Bayesian age model based on OSL and TL age estimates constrains the horizon containing the pierced *Nassarius* shells to between 73.4 ka and 91.5 ka with a most likely date of 82.5 ka.



Figure 6. *Nassarius gibbosulus* shell beads from the Mousterian levels of the Grotte des Pigeons, Taforalt, Morocco, and a present day shore (bottom right). (Modified after Bouzouggar *et al.* 2007, photo M. Vanhaeren and F. d'Errico)

Similar perforated *N. gibbosulus* are found at two other sites. The marine shells from Skhul, recovered during the McCown excavation at this site (McCown and Keith 1939) were recently located at the Department of Palaeontology, Natural History Museum, London and analysed by a multidisciplinary team led by two of us (Vanhaeren *et al.* 2006). Two *Nassarius gibbosulus* shells (Figure 7) bear perforations that could indicate suspension. To identify the layer from where these *Nassarius* come sediment matrix adhering to one of them and sediment samples from the three layers identified at the site (A, B1 and B2) were analyzed for mineralogy and chemical composition. Major and trace elements, as well as the hardness of the sediment adhering to the pierced shell indicate it comes from layer B1–2, dated by Electron Spin Resonance and Uranium-series analyses to between 100 ka and 135 ka years (Grün *et al.* 2005). The chronology of the single perforated specimen of *N. gibbosulus* from Oued Djebbana, is less secure. This shell comes from 1 meter thick archaeological layer situated under 3.9 meters of sterile alluvial deposits that contains typical Aterian and other Middle Palaeolithic tools. A single infinite conventional radiocarbon date of 35 ka B.P. (MC 657) is available for this site (Morel 1974, Vanhaeren *et al.* 2006).

The *Nassarius kraussianus* found in the MSA layers of Blombos Cave, Western Cape Province, South Africa are close in age to those discovered at the Grotte des Pigeons.

Excavations conducted at Blombos since 1992 (Henshilwood *et al.* 2001; Henshilwood 2007, 2009) have identified a stratigraphic sequence with, from the

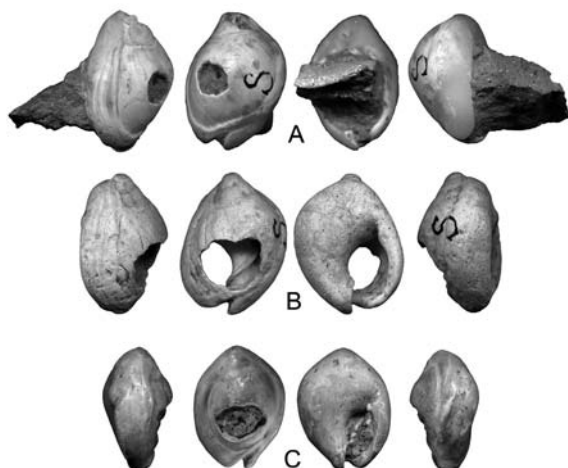


Figure 7. *Nassarius gibbosulus* shell beads from the Mousterian levels of Es-Skhul (A and B) and the Aterian levels of Oued Djebbana (C). Scale = 1 cm (modified after Vanhaeren *et al.* 2006, photo M. Vanhaeren and F. d'Errico)

top to the bottom, 80 centimetres of LSA deposit, an undisturbed 10–50 centimetres sterile aeolian dune sand, and a three main MSA phases (M1, M2, and M3) (Henshilwood *et al.* 2001, Henshilwood 2006, 2008). Principal markers of the M1 phase are bifacial foliate points, typical of the Still Bay technocomplex. Thirty-nine beads manufactured from *Nassarius kraussianus* gastropod shells (Figure 8) come from the upper MSA phase, M1, and two derive from the M2 phase (Henshilwood *et al.* 2004, d’Errico *et al.* 2005). The two shell beads found in M2 may be intrusive due to slumping of the deposits in the recovery area and probably originate from the overlying M1 phase. Thirty-three beads were found in six groups of two to twelve beads, each group being recovered in a single square or in two adjacent sub-squares. Morphometric, taphonomic, and microscopic analysis of *Blombos* shells and of biocoenoses and thanatocoenosis of modern shells of the same species indicate that their presence at the site cannot be due to natural processes (Henshilwood *et al.* 2004, d’Errico *et al.* 2005). Experimental reproduction indicates that the perforations were made by humans, probably with bone tools. Use-wear, recorded on the perforation edge, the outer lip, and the parietal wall of the aperture indicates the shells were strung and worn. The strings and the beads were deliberately stained with ochre judging from the remains of red pigments observed microscopically inside a number of shells.

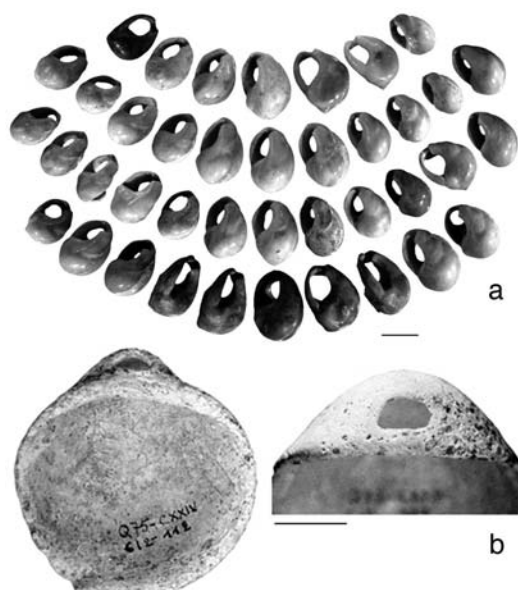


Figure 8. a: *Nassarius kraussianus* shell beads from Middle Stone Age layers of Blombos Cave dated at 75,000 years ago (modified after Henshilwood *et al.* 2004, photo F. d’Errico and C. Henshilwood). b: *Glycymeris* shell from the Mousterian layers of Qafzeh dated at ca 90 ka (after Taborin 2003). Scale bars = 1 cm.

Two other sites have yielded less compelling evidence for early bead use. Six *Afrolittorina africana*, three of which bear perforations (Figure 9) have been recently found in the ca. 70 ka Still Bay layers of Sibudu Cave, KwaZulu-Natal, South Africa, (d'Errico *et al.* 2008, Jacobs *et al.* 2008, Jacobs & Roberts 2008)). Taphonomic analysis of the specimens based on present day *Afrolittorina africana* biocoenoses, microscopic examination, morphometry, experimental perforation of modern shells, and a review of the natural agents that may accumulate marine shells at inland sites, indicate probable human involvement in the collection, transport, modification, and abandonment of *Afrolittorina africana* in Sibudu. Due to small sample size, however, this conclusion needs to be confirmed by future discoveries.

Other marine shells interpreted as beads come from the ca. 90 ka Mousterian levels XXI-XIV at Qafzeh Cave in Israel (Taborin 2003; Walter 2003; d'Errico & Vanhaeren 2007, Bar-Yosef Mayer *et al.* 2009). They consist of ten naturally perforated *Glycymeris insubrica* shells (Figure 8). Perforations present in few cases notches interpreted as damage produced by a string (Bar-Yosef Mayer *et al.* 2009). Naturally perforated still underscribed *Glycymeris* shells in Mousterian contexts are also present at Ras el Kelb Cave on the Lebanese coast and in Layer C of Sefunim Rockshelter in Israel (Lamdan 1984; Ronen 1984; Reese 1998). Unresolved evidence for a very ancient bead use come from a few Acheulean sites in Northern Europe. Excavations at a number of these sites at the end of the 19th century yielded fossil sponges belonging to the Late Cretaceous species *Porosphaera globularis*. The spherical morphology of these fossils, their small size, the presence of a natural complete perforation in their centre, and possible anthropogenic modifications have led some archaeologists to interpret them as the oldest known beads. Although our reappraisal of these objects (Rigaud *et al.* 2009) indicates that selection and modification by archaeologists rather than by Acheulean hominids is a more parsimonious explanation to account for the differences recorded between natural and archaeological samples, we are unable to completely rule out the bead hypothesis.

No convincing personal ornaments reliably dated to between ca. 70 ka and 40 ka are known from Africa and Eurasia. At around 40 ka, beads reappear, almost simultaneously, in Africa and the Near East, and appear for the first time in Europe and Australia. In Africa they take the form of ostrich eggshell beads and stone rings (McBrearty & Brooks 2000, d'Errico *et al.* 2005, d'Errico & Vanhaeren 2009). In the Near East perforated *N. gibbosulus*, the same species used at Skhul, Oued Djebbana and Grotte des Pigeons, is found at Uçagizli, South of Turkey (Kuhn *et al.* 2001) in layers dated to 41.4 ± 1.1 BP (AA 37625).

In Europe the earliest beadworks are associated with both Neanderthals and *H. sapiens* (d'Errico *et al.* 1998, 2003). They differ from their 100 ka -70 ka years antecedents in that they take the form of dozen of discrete types identifying regional patterns (Vanhaeren & d'Errico 2006).



Figure 9. a: *Afrolittorina africana* shells from the Still Bay layers of Sibudu Cave, KwaZulu-Natal; b: experimental double perforation of an *Afrolittorina africana* shell made to test the hypothesis that an archaeological specimen (bottom left) corresponds to a shell perforated after a first breakage (modified after d'Errico *et al.* 2007, photos d'Errico and Vanhaeren). Scale bars = 1 cm.

In Asia, eleven EUP sites from Siberia and in particular from the Altai and West Baikal have yielded personal ornaments (see d'Errico & Vanhaeren 2007 for a synthesis). Some of them were recovered from sites (Denisova, Podzvonkaya, Khotyk, Kara-Bom, Maloyalomanskaia) that are at least as old as the Aurignacian in Europe (37 ka) and could well be contemporaneous with the earliest European Upper Palaeolithic (40 ka). The repertoire of personal ornaments from these sites is varied (32 types recorded) and comparable to that observed at Aurignacian and Chatelperronian sites.

In South East Asia the oldest ornament is a perforated tiger shark tooth found in New Ireland, New Guinea at a site dated between 39.5 ka and 28 ka (Leavesley 2007). The earliest evidence for bead use in Australia comes from the site of Mandu Mandu, Cape Range of Western Australia, where 22 *Conus* sp. shell beads were recovered in a layer dated to ca. 32 ka BP (Morse 1993). In addition, ten Dentaliidae shell beads are reported from the 30 ka old layers of Riwi in the Kimberly of Western Australia, a site located 300 kilometres inland (Balme & Morse 2006).

Burial practices

Many researchers regard primary burials and, to a lesser extent, special treatment of the dead, as proof of symbolism. Burial of the dead in all traditional societies is a complex symbolic activity (Leroi-Gourhan 1964, Pearson 1999, Maureille 2004).

It is difficult to imagine that a human group could excavate a grave, position the corpse in the pit, and offer funerary goods with no form of verbal exchange. Language in such situations is necessary to organise the rituals, transmit customs, and express emotions. However, the intentional character and symbolic significance of burials prior to 30 ka, especially those of Neanderthals, has been the subject of intense debate over the past decades (Gargett 1989, 1999, Stringer & Gamble 1993, Noble & Davidson 1996, Mellars 1996a, b, 1998, Tillier *et al.* 1991). While Gargett (1989, 1999) is right in pointing out that the mere discovery of a buried body with preserved anatomical connections does not demonstrate the deliberate disposal of the dead, and that firm criteria must be applied to this field of study, a growing body of evidence contradicts the minimalist view of Middle Palaeolithic and, in particular, Neanderthal funerary practices (Tillier *et al.* 1991, Maureille & Van Peer 1998, Maureille 2004, Maureille & Vandermeersch 2007, Maureille & Tillier 2008). Most Neanderthal burials were excavated long ago and available reports often lack the information now required to assess the anthropogenic origin of the inhumation and whether items interpreted as grave goods were deliberately interred. Methodologically, however, criteria for a taphonomic analysis of human remains found in archaeological context have been developed in the last two decades (Duday & Masset 1987, Duday 1995, Courtaud & Duday 2008) and are increasingly applied to Palaeolithic case studies. Analytically, it is a fact that not less than 60 Middle Palaeolithic primary and possibly secondary burials are reported from Europe and the Near East (Figure 10), and that 41 of them belong to Neanderthals (Zilhão & Trinkaus 2003, Maureille 2004, Maureille & Vandermeersch 2007, Maureille & Tillier 2008). If the presence of human skeletons in caves was due to natural causes one might expect to find in the same deposits complete skeletons of other mammals, which is not the case (Belfer-Cohen & Hovers 1992). We should also find complete skeletons of hominines in older periods. Now, with the single exception of the Australopithecine Stw 573 (Clarke 1998), no complete hominin skeletons are known before the Middle Palaeolithic of Europe and the Near East. The skeleton of a pre-Neanderthal from Altamura, Italy (Delfino Pesce & Vacca 1994), whose death was probably caused by falling into a cracks, does not preserve any anatomical connection.

In a number of cases purposely dug pits are documented in a number of cases and the adult Neanderthal burial from Kebara shows the skull was removed well after the body was buried (Arensburg *et al.* 1985, Tillier *et al.* 1991, Pettitt 2002). Three burials, Qafzeh 11, Skhul V and Amud 7, all from the Near East, have yielded bone objects associated with the skeletons. The first two burials are attributed to *H. sapiens*, the last to Neanderthal (McCown 1937, Vandermeersch 1981, Rak *et al.* 1994). Although the sample size is too small to qualify these objects as “grave goods”, in all three cases the objects only consist of cephalic elements (maxilla,

mandible, antler), which may indicate a behavioural consistency. In other cases the presence of grave goods is more controversial due to the antiquity of the excavations. A case in point is that of La Ferrassie, where seven Neanderthal burials were excavated by Capitan and Peyrony, and subsequently by Delporte (Peyrony 1934, Heim 1976). According to the excavators one adult was buried with grave goods including stone tools, bone retouchers and a bone engraved with parallel lines; further burials include a 3-year-old child whose funeral pit was covered with cupules engraved rock slabs, and that of a foetus accompanied by a flint point and two scrapers. Funerary goods similar to those found at La Ferrassie are reported from La Chapelle-aux-Saints and Le Moustier in France, and in the Near East from Amud and Dederiyeh (Akazawa *et al.* 1995, Hovers *et al.* 2000).

TL dating of Tabun layer C at 170 ka (Mercier *et al.* 2000) seems to indicate that the Neanderthal burial found in this layer by Garrod predates the oldest known *H. sapiens* burial found at Qafzeh, or is, at least, contemporaneous with the latter (Grün & Stringer 2000). If this skeleton is indeed associated with level C and not the still undated level B, as recently envisaged by Bar-Yosef and Callander (1999), the new thermoluminescence dates would clearly contradict the interpretation of the Near East Neanderthal burials as the result of acculturation by neighbouring *H. sapiens*.



Figure 10. Middle Palaeolithic burials from Tabun (a), Qafzeh (b) and Kebara (c) caves, Israel, and hyoid bone from the Kebara burial (modified after Garrod and Bate 1937, Arensburg *et al.* 1985, Vandermeersch 1969, and by courtesy of B. Vandermeersch)

The earliest possible cases of special treatment of the dead are ambiguous. The cut-marks on the Bodo skull, Ethiopia, dated to ca 600 ka indicate defleshing with a stone tool suggesting either cannibalism or a postmortem ritualized treatment of the skull (White 1986). The team excavating the Sima de los Huesos site has proposed a symbolic interpretation for the accumulation of 32 to 28 pre-Neanderthals associated with a unique handaxe dated to between 300 ka and 400 ka (Carbonell *et al.* 2003).

More recently mortuary practices consisting of defleshing, removal of the mandible and curation of the skull have been proposed for the remains of three individuals found in layers dated to ca. 160 years ago at Herto, Ethiopia (Clark *et al.* 2003, White *et al.* 2003). Although similar interpretations have been proposed for Neanderthal remains from Europe, a number of instances have been formally discarded (White *et al.* 1991; Piperno & Giacobini 1992).

In sum, there is clear evidence that *both* Neanderthals *and* *H. sapiens* buried their dead but there are no clear-cut differences in their respective funerary behaviour. Burials in the Near East and elsewhere by both species differ from those of the Upper Palaeolithic of Europe in that they lack personal ornaments, bone tools and ochre. Many fragments of ochre bearing traces of use and ochred stone tools come from the Qafzeh layers which yielded the burials (Hovers *et al.* 2003), but none have been found in clear association with the skeletons. Apart from the burials from Nazlet Kather (Vermeersch *et al.* 1984, Crevecoeur 2006) and, perhaps, Qena, Egypt (Vermeersch *et al.* 1998) no burials of *H. sapiens* in Africa are known between 90 ka and c. 30 ka.

In Europe, little is known about funerary practices during the Middle – Upper Palaeolithic transition. Human bones attributed to the period between 40–30 ka BP are scarce and fragmentary (Gambier 1997, 2000, Gambier *et al.* 1990, Churchill & Smith 2000, Henry-Gambier *et al.* 2004). Available data on the concentration of Neanderthal human remains from Saint-Césaire (Charente) is insufficient to support the hypothesis of a primary or secondary burial (Vandermeersch 1993). Traditionally considered as Aurignacian, the famous Cro-Magnon burial is now attributed to the Early Gravettian on the basis of the Accelerator Mass Spectrometry date of a shell bead (Henry-Gambier 2002). No other primary burials are directly dated to more than 28 ka BP and the oldest known after this date are also attributed, like Cro Magnon, to the Gravettian (Pettitt & Bader 2000, Formicola *et al.* 2004, Pettitt *et al.* 2003, Svoboda *et al.* 2002, Sladek *et al.* 2000). The date of 34.8 ka – 30.4 ka BP (Oxa 7073) for Kostienki 1 (Richards *et al.* 2001) was obtained from an isolated long bone, not from a burial (Sinitsyn pers. comm.) The Early Aurignacian perforated human teeth from Brassempouy (Landes), La Combe (Dordogne), and Isturitz (Pyrénées-Atlantiques) are among the oldest Upper Palaeolithic human remains bearing traces of deliberate intervention that may be interpreted as evidence for complex funerary practices (Henry-Gambier *et al.* 2004). The size of these teeth

falls within the range of variation of Near Eastern Middle Palaeolithic and Gravettian *H. sapiens*, but also of the few known European Aurignacian Modern Humans and late Mousterian Neanderthals, which makes it difficult to draw conclusions concerning their taxonomic affiliation (but see Baley & Hublin 2005 for a different view). Elsewhere in Europe, Aurignacian remains are also fragmentary and equally undiagnostic. The best preserved remains, such as those from Mladeč, cannot be securely assigned to the Early Aurignacian. They were excavated with the same methods used to dig the Koneprusy burial, Moravia (Svoboda *et al.* 2002), now attributed to the Magdalenian. The recent dating of the Vogelherd remains to the Neolithic, previously considered Aurignacian (Conard *et al.* 2004), constitute further confirmation that ancient attributions cannot be trusted. Although discovered recently and directly dated to 34 ka, the skull and mandible from Oase, Romania, are still of unknown cultural attribution (Trinkaus *et al.* 2003). The three fragmented bones of a perinate, found at Le Piage (Champagne & Espitalié 1981), are the only known human remains associated with the so called “Archaic” Aurignacian. In summary, the anatomy of the makers of the ancient Aurignacian in Europe is at present too poorly known to firmly state that the early Aurignacian is the sole product of *H. sapiens*. It is also unclear whether there was genetic input from Neanderthals into the *H. sapiens* genome.

The earliest evidence for musical traditions

We have previously considered surviving evidence for acoustic tool-use in the Palaeolithic, and discussed its implications for the origins of music, musical tradition and related behaviours in *H. sapiens* (Lawson & d’Errico 2002, d’Errico *et al.* 2003, Lawson 2004, d’Errico & Lawson 2006). Foremost amongst the Upper Palaeolithic finds are collections of bird-bone and ivory pipes from Geissenklösterle, Hohle Fels, Vogelherd, Germany (Conard *et al.* 2009) and Isturitz, France (Figure 11). Whilst the evidence from Germany is the more closely dated, at around 33–36 ka, the latter is much larger, comprising more than twenty separate specimens from the same site; they range widely in date, from the Magdalenian back into the Aurignacian. In our analysis of the Isturitz pipes we paid special attention to three of the best-preserved and oldest pieces: one complete Gravettian pipe with four finger-holes, another almost complete Gravettian specimen, also with 4 holes, and the earliest Aurignacian piece. Much of this earliest pipe survives, including one finished end and no fewer than 3 finger-holes.

This analysis and comparison with the other extant specimens revealed surprisingly long-lived consistencies in form which are difficult to explain away as independent inventions. The Gravettian pipes have four finely-worked finger-holes,

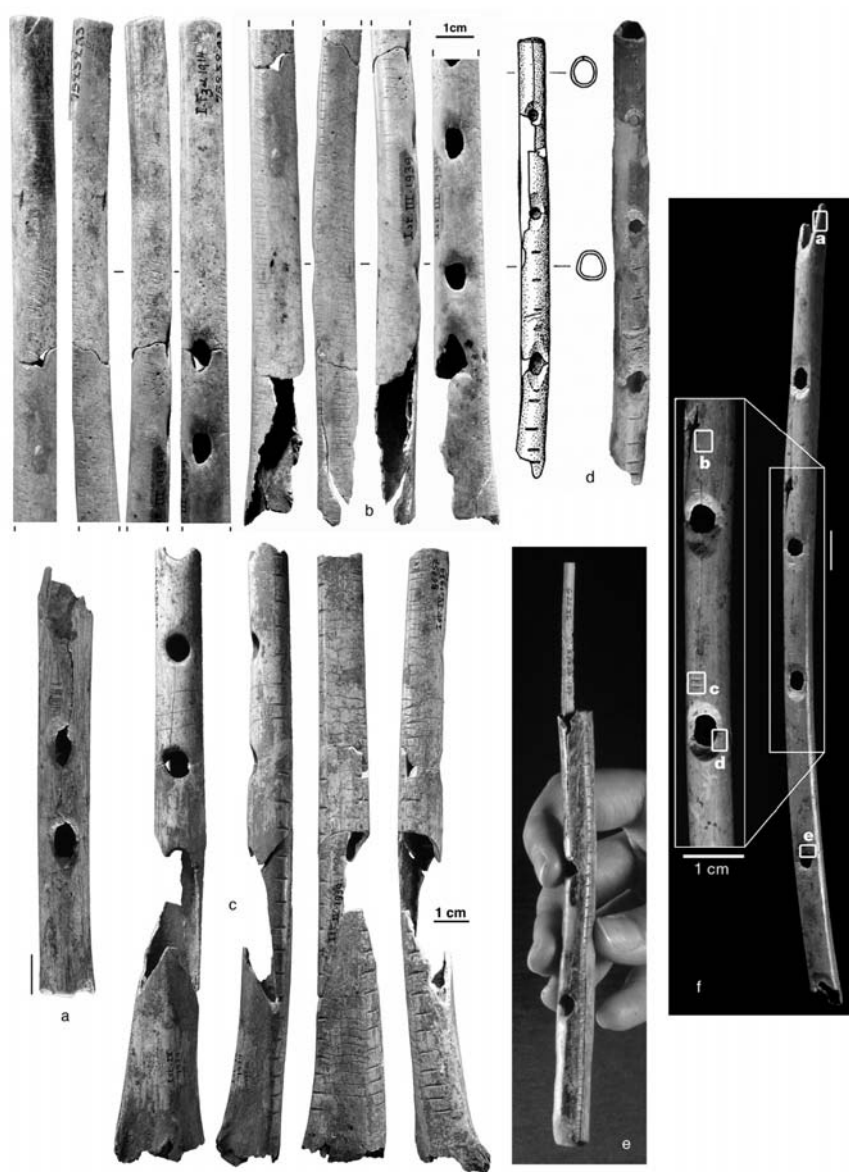


Figure 11. Upper Palaeolithic pipes made of raptor long bones from Isturitz, France, attributed to the Aurignacian (a) and the Gravettian (b-c) and Aurignacian pipes from Geissenklösterle (d-e), and Hohle Fels (f), Germany, made of swan long bone (d), mammoth ivory (e), and griffon vulture long bone (f). (Isturitz pipes modified after d'Errico *et al.* 2003; Geissenklösterle pipes by courtesy of Nick Conard, photo H. Jensen, copyright of the University of Tübingen, Hohle Fels modified after Conard *et al.* 2009), Scale bars = 1 cm.

equal in size, and arranged precisely in two pairs. Both were made from the ulna of a vulture. This evidence suggests a surprising agreement of purpose for pipes that are separated in time by hundreds or even thousands of years (radiocarbon dates for the Gravettian elsewhere indicate that this technocomplex spans from 28 ka to 21 ka BP). The shared finger-hole layout has no obvious parallel with bone pipes of later pre- and early historic music cultures, and seems to contrast sharply even with those remaining from Isturitz's own late Upper Palaeolithic (Magdalenian) levels. So it is all the more remarkable to find suggestions of such continuity with even earlier, Aurignacian practices. Isturitz's Aurignacian pipe resembles the Gravettian in several ways: firstly in choice of bone and competence of manufacture; secondly, exhibiting as many as three, probably more, well-formed finger-holes with obliquely angled finger-orientations; thirdly, in their embodiment of two different spacings, the wider space (between holes 2 and 3) bearing divisional incisions; and finally, the same plain, polished terminal, if the surviving end is indeed the instrument's proximal end. Finds from Geissenklösterle, Hohle Fels and Vogelherd (Hahn & Münzel 1995, Hein 1998, Conard *et al.* 2004; Conard *et al.* 2009), again exhibit competent workmanship and complexity of form; a similar preference by their makers for the wing-bones of large birds (swan, griffon vulture), and similar linear series of notation-like scratches. They again reveal between three and five well-formed finger-holes, with closely comparable platformed margins, spaced in some cases unequally in such a way that the larger gap again falls between holes 2 and 3, and again displays 'three-scratch' divisional markers. In terms of the diversity of other, especially later Upper Palaeolithic (e.g. Magdalenian) instrument forms, this is a *very* close match.

The most recent finds from Geissenklösterle, Hohle Fels and Vogelherd (Conard *et al.* 2004; 2009) reveals a similar form achieved using a strikingly different set of craft skills. Instead of naturally hollow bird-bone, pipes have first been shaped from solid ivory and split into two separate halves; each half has then been hollowed out and the whole subsequently bound and perhaps glued back together along a perfectly prepared, airtight seam. The technique is familiar enough from medieval and later music technologies (represented in Europe by the ivory *oliphant* horn, the wooden *krummhorn* and traditional wooden folk-trumpets of many kinds) but it is nevertheless a surprise, even a shock, to find such a labour-intensive strategy employed so long ago. Whoever did the work seems to have produced an instrument that closely resembles those made of bird-bone. We argue elsewhere (e.g. Lawson *in press*) that all such bone pipes, prehistoric or medieval, should likely be seen as the surviving tip of a musical iceberg, of which the broader mass would have been made up of plant-based instrument-making behaviours for which direct evidence could not be expected to survive. This new pipe may well offer a small measure of circumstantial support for that proposition. But most

importantly for our present argument it reveals form transcending not only time and space but also technology.

The sophistication of these pipes' various adaptations, both in their manufacture and their use, suggests that in cognitive terms they are, even at around 40 ka, already several stages removed from the earliest origins of instrumental musical expression, and earlier non-instrumental and dance forms. It is certainly difficult to see in them any indications that they represent the earliest attempts at piping. Acquisition of such skills must indeed have been remote. But how remote? Claims for earlier flutes-with-finger-holes have so far failed to stand up to close scrutiny. The so-called 'Neanderthal flute', a fragment of an immature cave-bear femur from Divje Babe II Cave, Slovenia (Turk 1997, Kunej & Turk 2008) has proved to be rather the result of natural processes (d'Errico *et al.* 1998, Chase & Nowell 1998, d'Errico & Lawson 2006). Similarly a long-bone shaft with a single perforation, found in the Middle Palaeolithic levels of Haua Fteah, Libya and published as a broken whistle by McBurney (1967) has proved to have similar origins (d'Errico & Villa 1997).

This does not mean that Neanderthals and African MSA people had no such musical behaviours, nor does it mean that we should abandon searching. We have to question whether we should be anticipating only 'pipes-with-finger-holes' at such early dates, or whether we should be looking for Pan-pipes – pipes made up of series of plain tubes *without* finger-holes – or other types of musical instruments (e.g. Zubrow & Blake 2006). Cross-cultural studies of modern hunter-gatherer instrument-making behaviours (e.g. Morley 2006) show how limited the chances are of these instruments surviving for thousands of years, to be correctly identified by archaeologists.

So if we accept ancient instruments (such as bone pipes) as proxies for aspects of ancient language, what might we infer about language origins from extant finds; and what kinds of finds should we seek in future to extend this perspective into earlier periods? At around 30 ka to 36 ka human visitors to sites such as Geissenklösterle and Isturitz not only *could* but *did* use sound in complex, articulated ways. They exhibit an individual ability to generate and distinguish discrete pitches (frequencies), and they demonstrate the intention to do so deliberately, in temporally organised sequences, in ways that are at least analogous with elements of the way we generate speech today. Without both of these abilities speech could not have developed. Such a date-range is still too recent to allow us to directly challenge the idea of a strict link between these innovations and biological change. However, the sophistication they exhibit at that first appearance in the archaeological record, *circa* 36 ka, could be consistent with a very lengthy antecedence. This is likely to have been incremental or discontinuous rather than linear, involving several major cognitive leaps. Of these the acquisition of the 'finger-hole

principle' to achieve serial tunings from one pipe, in place of an array of separately tuned single-tone pipes, would perhaps have represented the biggest step change. It remains in use wherever flutes and pipes – including some horns and trumpets – are played today.

Language, anatomy and encephalization

The lack of language, or at least the lack of sophisticated speech skills, has for long been one of the main reasons to consider Neanderthals as biologically handicapped compared with *H. sapiens* (Laitman *et al.* 1975, Binford 1989, Klein 1999, 2003, Mellars 1998, Noble & Davidson 1996, Lieberman 1975, 1989, 2006, 2007, Mithen 1996, Tattersal 1999, Lewis-Williams 2002). The Neanderthal larynx was considered similar to that of non-human primates or that of human children before the age of full vocalisation. The discovery of the hyoid bone with the Kebara 2 skeleton (Arensburg *et al.* 1989) and its strong similarity to that of *H. sapiens* suggests there were no major anatomical obstacles to Neanderthals having modern speech. The recent description of new pre-Neanderthal and Neanderthal hyoid bones confirms this first study (Rodriguez *et al.* 2003, Martinez *et al.* 2008). Analysis of two hyoid bodies from La Sima de los Huesos shows that their morphology is very similar to that of present day humans.

Although this view is not shared by all (e.g. Hublin 2005), recent research confirms there are no good grounds for different language abilities in Neanderthals and *H. sapiens* (Arensburg *et al.* 1990, Bresson 1992, Schepartz 1993, Tillier & Arensburg 2000, Boë *et al.* 2002, 2007). Lieberman, a critic of Neanderthal speech capabilities has accepted that the initial assessment of the La Chapelle-aux-Saints Neanderthal vocal tract is incorrect and that they were certainly able to speak. He suggests that “we can account for Neanderthal extinction without invoking any radical cognitive-linguistic distinction if we take account of their speech capabilities” (Lieberman 2001) and as far as the assessment of Neanderthal intelligence is concerned “only archaeological evidence, our window on hominid behaviour in the distant past, can resolve this question”.

As a confirmation of this cautious stand, a number of studies (e.g. Holloway 1985, 1995, Semendeferi 2001) have recently shown that palaeoneurology and functional anatomy are unable to detect significant differences between the neural development of Neanderthals and modern humans and to predict behavioural implications for the observed differences. A more recent study (Ponce de Léon *et al.* 2008) convincingly shows that Neanderthal brain size at birth was similar to that in recent *Homo sapiens*. Neanderthal brain growth rates during early infancy were higher, resulting in larger adult brain sizes but not in earlier completion of brain

growth. This suggests that Neanderthal growth rates were similar or even slower than recent *H. sapiens*.

A number of scholars believe that the “archaic” Neanderthal features must reflect inferior cognitive abilities (e.g. Tattersall 1999, Wynn & Coolidge 2004). Such a bias is clear in recent attempts to infer the timing of language origin from the phylogeny of encephalization. The relatively large brain size and estimated young age at death of a single *Homo erectus* child from Modjokerto, Indonesia, is used by Coqueugniot *et al.* (2004: 301) to claim that “*complex spoken language emerged relatively late in the course of human evolution*». However, implications of a more rapid brain growth for language acquisition are largely speculative, and the comparison the authors make with the brain size of chimpanzees of the same age may not be pertinent considering the crucial role played by the organization, rather than the volume, of the brain in determining cognitive capacities and language abilities. Little differences in cognitive capacities are detected between populations showing differences of up to 500 cm³ in their average brain size (e.g. Feuillet *et al.* 2007 for an extreme case).

Another morphological issue that has implications for the language debate is that of the evolution of the speech organs. Kay *et al.* (1998) have measured the size of the hypoglossal nerve groove, the main nerve controlling the tongue, and found that the size of that groove in the Neanderthal cranium was comparable to *H. sapiens*. They concluded that the movements of the tongue (i.e. numerous, varied and fast) were the same as in *H. sapiens*. A comparative study of the same groove in great apes as well as living and extinct hominids concludes that the size of the groove is not related to the size of the hypoglossal nerve itself (De Gusta *et al.* 1999).

A newer study, conducted by Jungers *et al.* (2003) demonstrated that humans possess significantly larger canals than any other species except gorillas, but there is considerable overlap with chimpanzees. *Australopithecus afarensis* is well within both human and gibbon ranges, as are the indices computed for selected representatives of fossil *Homo*. This suggests that the relative size of the hypoglossal canal is neither a reliable nor sufficient predictor of human-like speech capabilities.

In sum, we can now consider it a fact and not a hypothesis that the Neanderthal larynx was situated low in the throat, as in *Homo sapiens*. This has two important consequences: (1) the Neanderthal tongue was able to make a wide range of movements, because its base is located deep in the neck (this is a normal consequence of a bipedal stance – unlike all the other great apes), and (2) the Neanderthal soft palate was sufficiently separated from the epiglottis to allow laryngeal air into the buccal cavity. This implies that the bony anatomy of the Neanderthal did not prevent them from producing phonemes for articulate speech and there is, at present, no anatomical support for the hypothesis that there was a significant change in language abilities with the emergence of *H. sapiens*.

We may even wonder whether anatomical differences, if detected, may really tell us anything about differences in speech abilities. Considering the phylogenetic position of the Neanderthals in the human lineage, it is highly likely that their vocal tract was able to produce a wide range of phonemes. However, it is the articulation of these phonemes, rather than phonation itself, which is probably the crucial issue. Phonation is not peculiar to humans. It consists of producing an amplified sound when the breath crosses the larynx and the pharynx, which act as resonators. Study on the phonation capacities of fossil species in fact provides information on the pitch and quality of the sound rather than on speech. In contrast, the articulation of phonemes is a unique human capacity. It is the ability to utilise different phonemes that are easy to differentiate from each other, and to combine them to produce comprehensible sentences that communicate specific meaning.

Phoneme articulation in a linguistic fashion is fully related to the organisation of the brain and the nervous system. Unfortunately, information on the morphology and physiology of these organs in fossil populations is scant. We have few natural casts of human fossil brains (e.g. that of Ganocve for the Neanderthals) with which to study imprints of cerebral vessels (arteries, veins), nerves, and cerebral lobes on the cranial vault (Grimaud-Hervé 1997, Bookstein *et al.* 1999, Semendeferi & Damasio 2000). The study of aphasia (Jakobson 1980, Deacon 1997) helped locate the main brain areas involved in language production. Because the destruction of Broca and Wernicke areas of the cortex, both located in the left hemisphere, produces aphasia it was concluded that speech only involves the left hemisphere. We know now that the asymmetry of the two hemispheres is related both to language and to the preferential use of the right or the left hand. This asymmetry is already present in fossils of early *Homo* (Tobias 1987, 2005, Aiello & Dean 1990, Deacon 1992) and Neanderthals. The endocranial cast of the Amud 1 Neanderthal, for example, shows a markedly enlarged Broca's area on the left side of the brain (Ogawa *et al.* 1970), typical of modern right-handed people (Broca 1888, Falk 1987). Thus, the presence of Broca's and Wernicke's areas on a fossil skull is not sufficient evidence to argue for the existence of articulate speech. One might even speculate that structures morphologically similar to these areas might have existed in the past but were used by the brain for quite different functions.

We also know that a complex neurological circuit connects numerous motor, receptor and associative areas of the two hemispheres. In sum, while the presence of the Broca's and Wernicke's areas supports the existence of articulate speech, a lack of fossil evidence will continue to hinder definitive answers on this subject.

When language capacities are considered, the auditory system has the same importance as the vocal one. The recent research of Martinez *et al.* (2004) on the skeletal characteristics of the outer and middle ear underlines the fact that the auditory capacities of 5 pre-Neandertals (or *Homo heidelbergensis* for some

scholars) are supposed to be similar to those of living humans in the 3 to 5 kHz frequency range.

In order to speak, we must be able to control our breathing during speech. Following Maclarnon and Hewitt (2004), who examined thoracic vertebral canal size in fossil hominids and extant primates, *Homo ergaster* (based on the specimen KNM-WT 15000) does not have the same innervation of the thoracic muscles involved in the breathing control. As a consequence, they consider that «*Homo ergaster and Homo erectus could, at most, have produced short sequences of phonemes and morphemes. These would have provided sufficient sound combinations, for example, for an increased lexicon or vocabulary using the memory power of the enlarged brain, but communication would have remained limited, as can also be the case for modern humans which reduced breathing control as a result of severe spinal cord damage*».

With respect to their research one faces the same kind of limitations that we identified with Bruner *et al.* (2003) interpretations. *Homo ergaster*, *Homo erectus* and Neanderthals were not anatomically modern humans so it is normal to observe biological differences between them. Considering the chronological differences between these taxa especially. Differences in morphology cannot be equated directly with differences in cognition.

Concluding remarks

It is one of the ironies of the study of language origins that whilst *prima facie* archaeology and palaeoanthropology may not historically have been the best equipped disciplines to be asking what exactly language is and when it arose, it is beyond doubt that they hold in their hands the best, not to say the only, ancient material resources and information with which to address the latter question directly.

Of course, whether we accept manifestations of symbolic behaviour as “proxies” for language does not automatically tell us when and how language arose. It leaves open the question of whether language has spread from symbolism (Deacon 1997) or *vice versa* (Donald 1991). It does not explain how complex language needs to be to create and transmit a symbolic material culture, or what kinds of symbolic manifestation are more likely to reflect the presence, for example, of syntactical rules in that population's spoken language. Equally it does not provide us with any clue as to whether the switch to recursive syntactical language was a sudden event, as implied by Chomsky (1980) and developed by Bickerton (1981, 1990), or whether it was a gradual process (Pinker 2000, Jackendoff & Pinker 2005).

In order to test these alternative hypotheses we should first try to gain a better insight into the pertinence of different kinds of symbolic artefact as correlates of

language in general and syntax in particular. Correlates of syntactical functions, such as hierarchical organisation, *merge*, recursion and links between distant elements (Rizzi in press) are found in complex symbolic non-linguistic codes. The discovery in the archaeological record of symbolic artefacts displaying codes of such a complexity may identify in past populations cognitive abilities compatible with the development of syntactical language.

For the time being, and as provisional as it may seem to other disciplines engaged in the origins of language, archaeological and palaeoanthropological evidence seem to us to indicate that symbolic thinking and language are unlikely to be the result of any abrupt biological change but on the contrary that they had appeared gradually in the evolution of mankind. In particular, our review of the earliest evidence for symbolic material cultures does not support simple versions either of the “Human Revolution” model or the “Out-of-Africa” model for the emergence of modern behaviour and language.

The former scenario is contradicted by a growing body of evidence indicating that in Africa, the Near East and, to a lesser extent, in Europe human populations were involved in symbolic activities well before both the arrival of *H. sapiens* in Europe and the 50 ka-year-old rapid neural mutation which, according to some authors, would have qualitatively changed human cognition. The impressive record from the Blombos and the Sibudu sites, and the recent discovery of shell beads at ancient sites from North Africa and the Near East are paradigmatic in this respect.

On the other hand, the unequivocal presence of funerary practices and pigment use at a number of Neanderthal sites from Europe and the Near East contradicts a strict link between these innovations and the origin of our species in Africa. The fact that such behaviours occur in Europe well before any contact with *H. sapiens* further strengthens this observation.

Potential discrepancies also exist between the Out-of-Africa model and the first use of pigment in Africa. Early pigment use is reported at African sites such as Kapthurin and Twin Rivers that are apparently much older than the estimate that geneticists propose for the origin of our species (Ingman *et al.* 2000), and older than the date of the oldest known African fossils attributed to modern humans (Stringer 2002, Clark *et al.* 2003, McDougall *et al.* 2005). This indicates to us that it was archaic *Homo sapiens* and not modern humans that may have been responsible for the first use of pigments in Africa, thus breaking the functional link established by the Out-of-Africa scenario between biological and cognitive change. The publication of a detailed description of the alleged pigments from the Kapthurin formation is needed to test this hypothesis.

The production and use of a varied repertoire of personal ornaments by Neanderthals at the end of their evolutionary trajectory also contradicts the second model since it demonstrates that this alleged hallmark of modernity was accessible

to other fossil species. The presence of personal ornaments at Chatelperronian and Uluzzian sites has been interpreted as the consequence of acculturation of local Neanderthals by incoming Aurignacians (Hublin *et al.* 1996, Mellars 2005, Gravina *et al.* 2005, Mellars *et al.* 2007), to independent cultural evolution of Neanderthals before the spread of the Aurignacian (d'Errico *et al.* 1998, Zilhao & d'Errico 1999, 2003a,b, Zilhao 2001, 2006, Zilhao *et al.* 2006, 2008), or emerging from cross-cultural fertilization of Chatelperronian/Uluzzian Neanderthals and Aurignacian *H. sapiens* (d'Errico *et al.* 1998, White 2001). Even if it was demonstrated, which is far from being the case, that the use of personal ornaments by Neanderthals was the result of cultural contact this could in fact reinforce rather than dismiss the modern character of their cognition. This is because it could equally be consistent with an ability to incorporate external stimuli and reshape those influences in order to make them an integral part of their culture. Such abilities may be observed in many historical instances among modern human populations.

From their variety in morphology, colour, raw material, perforation and shaping techniques as well as geographic variation and association, the personal ornaments found at Aurignacian sites fit perfectly the interpretation of their reflecting complex codes, conceived to convey meaning to the members of the same or neighbouring groups by means of a shared symbolic language (Vanhaeren & d'Errico 2006, d'Errico & Vanhaeren 2009). Such a complexity, which would represent a comparable if not higher level than that observed in some historically known traditional societies, seems to us to imply language abilities equivalent to ours. If we accept this we must also, by the same token, grant similar language abilities to Chatelperronian Neanderthals. At least fifteen different personal ornament types, produced with different raw materials (teeth, fossils, ivory, bone) and manufacturing techniques are attested at Chatelperronian sites. Nine types are found at the Grotte du Renne alone. The number of bead types found at Aurignacian sites varies between one and forty (Vanhaeren & d'Errico 2006). The latter figure, however, is only found at the multistratified site of Mochi, Italy. Most of the other sites have yielded fewer bead types than the Chatelperronian levels of the Grotte du Renne. This suggests that the symbolic codes embodied in personal ornaments by late Neanderthals were of a complexity comparable to those used by Aurignacians when using the same media. The inescapable corollary of this observation is that Neanderthals must have had a communication system at least equivalent to the one we can infer for Aurignacian moderns.

Another observation that should give pause for thought before accepting the original "Out-of-Africa" scenario (McBrearty & Brooks 2000) is that no continuity nor real accretion of innovations has yet been observed in Africa. After their first occurrence in the archaeological record, at Blombos and possibly Sibudu, beads are not found in Southern Africa at sites attributed to the Howieson Poort, the

archaeological culture that follows stratigraphically (Wadley 2006, Rigaud *et al.* 2006) the cultural entity (Still Bay) to which shell beads are associated at Blombos Cave and Sibudu. They reappear in the same region 35 ka later in the form of ostrich eggshell beads. A similar picture can be seen in North Africa and the Near East. At Taforalt shell beads are found in layers dating from around 82 ka or more years ago but are absent from younger North African sites until the Upper Palaeolithic. New unpublished evidence for shell bead use at Aterian sites from Morocco support this pattern. The same applies to the Near East where perforated shells occur sporadically at old sites but are absent afterwards, and only reappear with the Ahmarian, some 50 ka later. This pattern may in part be attributable to a lack of investigation. However, a large number of sites younger than those that have yielded personal ornaments have been excavated in North Africa and the Near East, and a comparable number in Southern Africa. Against this background it would be reasonable to expect that while more ancient sites with shell beads will be identified in future they will not easily fill the gap.

The evidence presented here strengthens the view that such ornaments become obsolete after 70 ka in Africa and the Near East. This observation, and the different appearance of ornaments used after 40 ka in Eurasia, Africa and Australia support the view that major discontinuities in cultural transmission of innovations occurred 70 ka ago. Other cultural innovations arisen during the last interglacial such as the production of formal bone tools (d'Errico and Henshilwood 2007), abstract engravings (Henshilwood *et al.* 2001; 2009), bifacial shaping of projectile points (Jacobs *et al.* 2008) all appear to suffer the same fate before reappearing in different forms 10 ka later or more. Although the reasons and mechanisms behind this discontinuity are at present unknown, the long and intense climatic deterioration that characterizes in both hemispheres the MIS 4 (Lambert *et al.* 2008; Jouzel *et al.* 2007), between 73 and 60 ka ago, may be a good candidate to explain the rise and fall of these innovations. Jacobs and collaborators (2008) identify a clear gap corresponding to the MIS4, in the archaeological record of Southern Africa but discard the climatic hypothesis. They may well be wrong and MIS 4 has been proposed by a number of authors as a good candidate to explain the rise and fall of these innovations (Lahr and Foley 1998; Henshilwood 2008; Shennan 2001; Powell *et al.* 2009). In particular, by applying computer simulation Powell *et al.* (2009) have shown that patterns of innovation during the last climatic cycle is accounted for by expected population size and contact range variation. In such a context, marine shell beads may have been instrumental in creating and maintaining exchange networks between coastal and inland areas during a period of demographic growth towards the end of MIS 5.

Research on the relationship between cultural adaptation, geographic extent of cultural traditions and ecological constraints has shown that cultural diversity

decreases in areas with high ecological risk such as cold and arid environments (Nettle 1998; Collard and Foley 2002; Rockman and Steele 2003). This is explained by a need to creating long-distance social net-works, necessary for the exchange of information, goods and the exploitation of elusive resources. These areas are also those in which hunter-gatherers more frequently develop specialized technologies for large game hunting as bone and foliate projectile points (Binford 2001). Marine shell beads may have been instrumental in creating and maintaining exchange networks between coastal and inland areas and ensuring cultural cohesiveness in a period of climatic deterioration within MIS 5 or at the end of it. The arrival of harsher conditions of MIS 4 and their impact on human demography may have broken this mesh, depopulated some areas, isolated hunter-gatherer communities, and made the use of these innovations no longer advantageous. This scenario need to be tested. However, evidence presented here demonstrates that a number of innovations arose during the second half of the last interglacial and disappeared soon after its end. The emergence of such innovations represents one of the most fascinating cultural experiments of human history. This phenomenon, which may shed new light on the mechanisms of cultural transmission of early *Homo sapiens* populations, needs to be fully explored.

To conclude, the evidence and contradictions highlighted above suggest that the cognitive prerequisites of modern human behaviour were in place earlier in time amongst both “archaic” and “modern” populations, and we must invoke, among other factors, historical contingencies triggered by climatic and demographic factors rather than a single speciation event or mutation to explain the emergence, disappearance, and re-emergence in the archaeological records of symbolic traditions. According to Hovers and Belfer-Cohen (2005) Middle Palaeolithic populations underwent recurrent demographic crashes which both reduced their capacity to store knowledge and obliged them to “re-invent” again the same or similar innovations. Were Middle Palaeolithic societies more vulnerable to environmental changes due to their social systems or ways of transmitting knowledge? Complex languages probably emerged before the first appearance of complex symbolic artefacts in the archaeological record. This may have occurred in Africa, in conjunction with but certainly not as a simple result of, the origin of our species in that continent, or independently at different times and places, and in different forms, amongst different fossil populations. The latter scenario is consistent with recent genetic evidence (Krause *et al.* 2007) indicating that a critical gene known to underlie speech – known as the FOXP2 gene – was present in the Neanderthal genome and that its appearance predates the common ancestor, dated to around 300–400,000 years ago, of modern humans and the Neanderthals. The presence of FOXP2 in Neanderthals suggests that this speech demand was present when our two lineages shared a common ancestor.

Acknowledgements

Financial support for this project came from the OMLL programme of the European Science Foundation, the Programme Origine II of the Aquitaine Region, the Wenner-Gren Foundation, the French Ministry of Research, the PROTEA French-South Africa exchange programme, and postdoctoral grants from the Centre National de la Recherche Scientifique and the Fyssen Foundation to M.V. We thank Andrea Sajnerova-Duskova for granting us permission to use the photo of Mousterian pigments from Beçov.

References

- Aiello, Leslie C. 1998. "The foundation of human language." In *The origin and diversification of language* [Memoirs 24], Nina G. Jablonski and Leslie C. Aiello (eds), 21–34. San Francisco (CA): California Academy of Science.
- Aiello, Leslie C. and Cristopher Dean. 1990. *An Introduction to Human Evolutionary Anatomy*. New York: Academic Press.
- Akazawa, Takeru, Muhesen, Sultan, Dodo, Yukio, Kondo, Osamu, Mizoguchi, Yuji, Abe, Y., Nishiaki, Yoshihiro, Ohta, S., Oguchi, T. and Haydal, J. 1995. "Neanderthal infant burial from the Dederiyeh Cave in Syria." *Paléorient* 21: 77–86.
- Anikovich, Mikhail V., Sinitsyn, Andrei A., Hoffecker, John F., Holliday, Vance T., Popov, V. V., Lisitsyn S. N., Forman, S. L., Levkovskaya, G. M., Pospelova, G. A., Kuz'mina, I. E., Burova, N. D., Goldberg, Paul, Macphail R. I., Giaccio, B. and Praslov, Nicolai D. 2007. "Early Upper Paleolithic in Eastern Europe and Implications for the Dispersal of Modern Humans." *Science* 315: 223–226.
- Arensburg, Baruch, Bar Yosef, Ofer, Chech, M., Goldberg, Paul, Laville, Henri, Meignen, Liliane, Rak, Y., Tchernov, E., Tillier, Anne-Marie and Vandermeersch, Bernard. 1985. "Une sépulture néandertalienne dans la grotte de Kébara (Israël)." *Comptes Rendus de l'Académie des Sciences de Paris* 300(2): 227–230.
- Arensburg, Baruch, Schepartz, L. A., Tillier, Anne-Marie, Vandermeersch, Bernard and Rak, Y. 1990. "A reappraisal of the anatomical basis for speech in Middle Palaeolithic hominids." *American Journal of Physical Anthropology* 83: 137–146.
- Arensburg, Baruch, Tillier, Anne-Marie, Vandermeersch, Bernard, Dудay, Henri, Schepartz, L. A. and Rak, Y. 1989. "A Middle Palaeolithic human hyoid bone." *Nature* 338: 758–760.
- Backwell, Lucinda and d'Errico, Francesco. 2005. "The origin of bone tool technology and the identification of early hominid cultural traditions." In *From tools to symbols. From Early Hominids to Modern Humans*, Francesco d'Errico and Lucinda Backwell (eds), 238–275. Johannesburg: Wits University Press.
- Bailey, Shara E. and Hublin, Jean-Jacques. 2005. "Who made the Early Aurignacian? A reconsideration of the Brassempouy dental remains." *Bulletin et Mémoires de la Société d'Anthropologie de Paris* 17: 115–121.
- Bailey, Shara E. and Hublin, Jean-Jacques. 2006. "Dental remains from the Grotte du Renne at Arcy-sur-Cure (Yonne)." *Journal of Human Evolution* 50: 485–508.

- Balme, Jane and Morse, Kate. 2006. "Shell Beads and Social Behaviour in Pleistocene Australia." *Antiquity* 80 (310): 799–811.
- Barbujani, Guido. 2003. "Genetics and the population history of Europe." In *Nature Encyclopedia of the Human Genome*, David Cooper (ed), 1–5. London: Nature Publishing Group.
- Barham, Lawrence S. 1998. "Possible early pigment use in south-central Africa." *Current Anthropology* 39: 703–710.
- Barham, Lawrence S. 2002. "Systematic pigment use in the Middle Pleistocene of south-central Africa." *Current Anthropology* 31: 181–190.
- Bar-Yosef, Ofer. 1998. "On the nature of transitions: the Middle to Upper Palaeolithic and Neolithic Revolution." *Cambridge Archaeological Journal* 8: 141–163.
- Bar-Yosef, Ofer. 2002. "The Upper Palaeolithic Revolution." *Annual Review of Anthropology* 31: 363–393.
- Bar-Yosef, Ofer and Callander, Jane. 1999. "The woman from Tabun: Garrod's doubts in historical perspective." *Journal of Human Evolution* 37: 879–885.
- Bar-Yosef Mayer, Daniela, Vandermeersch, Bernard, Bar-Yosef, Ofer. 2009. "Shells and ochre in Middle Paleolithic Qafzeh Cave, Israel: indications for modern behaviour." *Journal of Human Evolution* 56: 307–314.
- Bednarik, Robert G. 2003. "A figurine from the African Acheulian." *Current Anthropology* 44 (3): 405–413.
- Behm-Blancke, Günter. 1983. "Altpaläolithische Gravuren von Bilzingsleben." *Ethnographisch-Archäologisch Zeitschrift* 24: 304–20.
- Belfer-Cohen, Anna and Hovers, Erella. 1992. In the Eye of the Beholder: Mousterian and Natufian burials in the Levant. *Current Anthropology* 34: 463–71.
- Berlin, Brent and Kay, Paul. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley (CA): University of California Press.
- Bickerton, Derek. 1981. *Roots of Language*. Ann Arbor: Karoma Publishers.
- Bickerton, Derek. 1990. *Language and Species*. Chicago: University of Chicago Press.
- Bickerton, Derek. 2003. "Symbol and structure: a comprehensive framework for language evolution." In *Language Evolution*, Christiansen Morton H. and Simon Kirby (eds), 77–93. Oxford: Oxford University Press.
- Binford, Louis R. 1983. *In Pursuit of the Past: Decoding the Archaeological Record*. London: Thames and Hudson.
- Binford, Lewis R. 1989. "Isolating the transition to cultural adaptations: an organizational approach." In *The emergence of modern humans: biocultural adaptations in the later Pleistocene*, Eric Trinkaus (ed), 18–41. Cambridge: Cambridge University Press.
- Binford, Lewis R. 2001. *Constructing Frames of Reference: An Analytical Method for Archaeological Theory Building Using Hunter-Gatherer and Environmental Data Sets* (University of California Press, Berkeley).
- Boë, Louis-Jean, Heim, Jean-Louis, Honda, Kiyoshi and Maeda, Shinji. 2002. "The potential Neandertal vowel space was as large as that of modern humans." *Journal of Phonetics* 30: 465–484.
- Boë, Louis-Jean, Heim, Jean-Louis, Honda, Kiyoshi S, Maeda, Shinji, Badin, Pierre and Abry, Christian. 2007. "The vocal tract of newborn humans and Neanderthals: Acoustic capabilities and consequences for the debate on the origin of language. A reply to Lieberman (2007a)." *Journal of Phonetics* 35: 564–581.
- Bookstein, Fred, Schäfer, Katrin, Prossinger, Hermann, Seidler, Horst, Fieder, Martin, Stringer Chris, Weber, Gerhard W., Arsuaga, Juan Luis, Slice Dennis E. F., Rohlf, James, Recheis,

- Wolfgang, Mariam, A. J. and Marcus, L. F. 1999. "Comparing frontal cranial profiles in archaic and modern *Homo* by morphometric analysis." *The Anatomical Record Part B: The New Anatomist* 257 (6): 217–224.
- Bordes, François H. 1952. "Sur l'usage probable de la peinture corporelle dans certaines tribus Moustériennes." *Bulletin de la Société Préhistorique Française* 49: 169–171.
- Bordes, François H. 1972. *A Tale of Two Caves*. New York (NY): Harper and Row Publishers.
- Botha, Rudolf. 2008. "Prehistoric shell beads as a window on language evolution." *Language & Communication* 28 (3): 197–212.
- Bouzouggar, Abdeljalil, Barton, Nick, Vanhaeren, Marian, d'Errico, Francesco, Collcutt, Simon, Higham, Tom, Hodge, Edward, Parfitt, Simon, Rhodes, Edward, Schwenninger, Jean-Luc, Stringer, Chris, Turner, Elaine, Ward, Steven, Moutmir, Abdelkrim and Stambouli, Abdelhamid. 2007. "82,000-year-old shell beads from North Africa and implications for the origins of modern human behaviour." *Proceedings of the National Academy of Sciences* 104: 9964–9969.
- Bowler, James M., Johnston, Harvey, Olley, Jon M., Prescott, John R., Roberts Richard G., Shawcross Wilfred and Spooner, Nigel A. 2003. "New ages for human occupation and climatic change at Lake Mungo, Australia." *Nature* 421: 837–840.
- Bresson, Frederique. 1992. "Aptitude au langage chez les Néandertaliens: apport d'une approche pluridisciplinaire." *Bulletin et Mémoire de la Société d'Anthropologie de Paris* 4, 33–51.
- Broca, Paul. 1888. *Recherches sur la localisation de la faculté de langage articulé*. Paris: Mémoire d'Anthropologie 5.
- Brown, Steven, 2000. "The "Musilanguage" Model of Music Evolution." In *The Origins of Music*, Nils L. Wallin, Björn Merker and Steven Brown (eds), 271–300. Cambridge (MA): MIT Press.
- Bruner, Emiliano, Manzi, Giorgio and Arsuaga, Juan Luis. 2003. "Encephalization and allometric trajectories in the genus *Homo*: evidence from the Neandertal and modern lineages." *Proceedings of the National Academy of Science* 100: 15335–15340.
- Byrne, Richard. 1995. *The thinking Ape*. Oxford: Oxford University Press.
- Cann, Rebecca L., Stoneking, Mark and Wilson, Allan C. 1987. "Mitochondrial DNA and human evolution." *Nature* 356: 389–90.
- Carbonell Eudald, Mosquera, Marina, Ollé, Andreu, Rodriguez, Xosé-Pedro, Sala, Robert, Vergès, Josep Maria, Arsuaga, Juan Luiz, and Bermudez de Castro, José-Maria. 2003. "Les premiers comportements funéraires auraient-ils pris place à Atapuerca, il y a 350.000 ans?" *L'Anthropologie* 107: 1–14.
- Cavalli-Sforza, Luigi Luca, Menozzi, Paolo, Piazza, Alberto. 1994. *The History and Geography of Human Genes*. Princeton: Princeton University Press.
- Champagne, F. & R. Espitalié, 1981. Le Piage, site préhistorique du Lot. *Mémoires Société Préhistorique Française* 15, 205.
- Chase, Philip G. 2006. *The Emergence of Culture. The Evolution of a Uniquely Human Way of Life*. New York: Springer.
- Chase, Philip G. and Nowell, April. 1998. "Taphonomy of a suggested Middle Paleolithic bone flute from Slovenia." *Current Anthropology* 39: 184–194.
- Chase, Philip G. and Dibble, Harold. 1987. "Middle Palaeolithic symbolism: a review of current evidence and interpretations." *Journal of Anthropological Archaeology* 6: 263–296.
- Chomsky, Noam. 1980. *Rules and Representations*. New York (NY): Columbia University Press.
- Churchill, Steven E. and Smith, Fred H. 2000. "Makers of the Early Aurignacian of Europe." *Yearbook of Physical Anthropology* 43: 61–115.

- Clark, J. Desmond, Beyene, Yonas, Wolde, Gabriel G., Hart, William K., Renne, Paul R., Gilbert, Henry, Defleur, Alban, Suwa, Gen, Katoh, Shigehiro, Ludwig, Kenneth R., Boisserie, Jean-Renaud, Asfaw, Berhane and White, Tim D. 2003. "Stratigraphic, chronological and behavioural contexts of Pleistocene *Homo sapiens* from Middle Awash, Ethiopia." *Nature* 432: 747–752.
- Clarke, Ronald J. 1998. "First ever discovery of a well-preserved skull and associated skeleton of *Australopithecus*." *South African Journal of Science* 94: 460–463.
- Clottes, Jean. 2003. *Return to Chauvet Cave: Excavating the Birthplace of Art: the First Full Report*. London: Thames & Hudson.
- Collard, Ian F. and Foley, Robert A. 2002. "Latitudinal patterns and environmental determinants of recent human cultural diversity: do humans follow biogeographical rules?" *Evolutionary Ecology Research* 4: 371–383.
- Conard, Nick J. 2005. "An overview of the patterns of behavioral change in Africa and Eurasia during the Middle and Late Pleistocene." In *From Tools to Symbols from Early Hominids to Humans*, Francesco d'Errico and Lucinda Backwell (eds), 220–235. Johannesburg: Wits University Press.
- Conard, Nick J. and Bolus, Michael. 2003. "Radiocarbon dating and the appearance of modern humans and timing of cultural innovations in Europe: new results and new challenges." *Journal of Human Evolution* 44: 332–371.
- Conard, Nick J., Grootes Pieter M. and Smith, Fred H. 2004. "Unexpectedly recent dates for human remains from Vogelherd." *Nature* 430: 198–201.
- Conard, Nick J., Malina, Maria, Münzel, Susanne C. 2009. New flutes document the earliest musical tradition in southwestern Germany. *Nature* doi:10.1038/nature08169.
- Conard, Nick J., Malina, Maria, Münzel, Susanne C. and Seeberger, Friedrich. 2004. "Eine Mammutfelsenflöte aus dem Aurignacien des Geissenklösterle: neue Belege für eine musikalische Tradition im Frühen Jungpaläolithikum auf der Schwäbischen Alb." *Archäologisches Korrespondenzblatt* 34: 447–462.
- Coqueugnot, Hélène, Hublin, Jean-Jacques, Veillon, F., Houët, Francis and Jacob, T. 2004. "Early brain growth in *Homo erectus* and implications for cognitive ability." *Nature* 431, 299–302.
- Courtaud, Patrice and Duda, Henri. 2008. "Qu'est-ce qu'une sépulture, comment la reconnaître?" In *Première humanité, gestes funéraires des Néandertaliens*, Bernard Vandermeersch, Jean-Jacques Cleyet-Merle, Jacques Jaubert, Bruno Maureille and Alain Turq (eds), 15–24. Paris: Réunion des Musées Nationaux.
- Crémades, Michèle, Laville, Henri, Sirakov, Nikolai and Kozłowski, Janusz K. 1995. "Une pierre gravée de 50 000 ans B.P. dans les Balkans." *Paléo* 7: 201–9.
- Crevecoeur, Isabelle. 2006. "Etude anthropologique des restes humains de Nazlet Khater (Paléolithique supérieur, Egypte)." PhD thesis, University of Bordeaux 1.
- Cross, Ian. 1999. "Is music the most important thing we ever did? Music, development and evolution." In *Music, Mind and Science*, Suk Won Yi. (ed), 10–39. Seoul: Seoul National University Press.
- d'Errico, Francesco. 1991. "Carnivore traces or Mousterian skiffle?" *Rock Art Research* 8: 61–63.
- d'Errico, Francesco. 2003. "The invisible frontier. A multiple species model for the origin of behavioral modernity." *Evolutionary Anthropology* 12: 188–202.
- d'Errico, Francesco. 2009. "Le Rouge et le Noir. Implications of early pigment use in Africa, the Near East, and Europe for the origin of cultural modernity." In *Current themes in Middle Stone Age research*, Marlize Lombard, Ester A. Huysen, Christine Sievers, (eds), 157–174. Goodwin Series Volume in honour of Lyn Wadley: South African Archaeological Society 10.

- d'Errico, Francesco and Lawson, Graeme. 2006. "The Sound Paradox. How to assess the acoustic significance of archaeological evidence?" In *Archaeoacoustics*, Chris Scarre and Lawson Graeme (eds), 41–57. Cambridge: McDonald Institute Monographs.
- d'Errico, Francesco and Nowell, April. 2000. "A new look at the Berekhat Ram figurine: implications for the origins of symbolism." *Cambridge Archaeological Journal* 10: 123–167.
- d'Errico, Francesco and Henshilwood, Christopher. 2007. "Additional evidence for bone technology in the southern African Middle Stone Age." *Journal of Human Evolution* 52: 142–163.
- d'Errico, Francesco and Vanhaeren, Marian. 2007. "Evolution or Revolution? New Evidence for the Origin of Symbolic Behaviour In and Out of Africa." In *Rethinking the Human Revolution*, Mellars, Paul, Boyle, Kate, Bar-Yosef, Ofer and Stringer Chris, (eds), 275–286. Cambridge: McDonald Institute Monographs.
- d'Errico, Francesco and Vanhaeren, Marian, 2009. "Earliest personal ornaments and their significance for the origin of language debate" In *The Cradle of Human Language*, eds. Rudolf Botha and Chris Knight, 16–40. Oxford: Oxford University Press.
- d'Errico, Francesco and Villa, Paola. 1997. "Holes and grooves: the contribution of microscopy and taphonomy to the problem of art origins." *Journal of Human Evolution* 33: 1–31.
- d'Errico, Francesco, Villa, Paola, Pinto, Anna and Idarraga Rosa. 1998. "A Middle Paleolithic origin of music? Using cave bear bone accumulations to assess the Divje Babe I bone «flute»." *Antiquity* 72: 65–79.
- d'Errico, Francesco, Henshilwood Christopher, Nilssen, Peter. 2001. "An engraved bone fragment from 70,000-year-old Middle Stone Age levels at Blombos Cave, South Africa: implications for the origin of symbolism and language." *Antiquity* 75: 309–318.
- d'Errico, Francesco Henshilwood, Christopher, Lawson, Graeme, Vanhaeren, Marian, Soressi, Marie, Bresson, Frederique, Tillier, Anne-Marie, Maureille, Bruno, Nowell, April, Backwell, Lucinda, Lakarra, Joseba A. and Julien, Michèle. 2003. "The search for the origin of symbolism, music and language: a multidisciplinary endeavour." *Journal of World Prehistory* 17: 1–70.
- d'Errico, Francesco, Henshilwood, Christopher, Vanhaeren, Marian and van Niekerk, Karen. 2005. "Nassarius kraussianus shell beads from Blombos Cave: evidence for symbolic behaviour in the Middle Stone Age." *Journal of Human Evolution* 48: 3–24.
- d'Errico, Francesco, Zilhão, João, Baffier, Dominique, Julien, Michèle and Pelegrin Jacques. 1998. "Neandertal acculturation in western Europe? A critical review of the evidence and its interpretation." *Current Anthropology* 39: 1–44.
- d'Errico, Francesco, Vanhaeren, Marian and Wadley, Lyn. 2008. "Possible shell beads from the Middle Stone Age layers of Sibudu Cave, South Africa." *Journal of Archaeological Science* 35: 2675–2685.
- De Lumley, Henri. 1969. « A Palaeolithic camp site at Nice." *Scientific American* 220: 42–50.
- De Waal, Frans B.M. and Pollick, Amy S. 2007. "Ape gestures and language evolution." *Proceedings of the National Academy of Sciences* 104: 8184–89.
- Deacon, Terrance, 1992. "Impressions of ancestral brains." In *Encyclopedia of Human Evolution*, Stephen Jones, Robert Martin and David Pilbeam (eds), 116–117. Cambridge: Cambridge University Press.
- Deacon, Terrance W. 1997. *The Symbolic Species: The Co-Evolution of Language and the Brain*. New York: Norton.
- DeGusta, David W., Henry Gilbert W and Turner, Scott P. 1999. "Hypoglossal canal size and hominid speech." *Proceedings of the National Academy of Science* 94: 1800–1804.
- Delfino Pesce, Vittorio and Vacca, Eligio. 1994. "Report of an archaic human skeleton discovered at altamura (Bari), in the "Lamalunga" district." *Human Evolution* 9: 1–9.

- Demars, Pierre-Yves. 1992. "Les colorants dans le Moustérien du Périgord. L'apport des fouilles de F. Bordes." *Préhistoire Ariégeoise* 47: 185–194.
- Dibble, Harold L., 1989. « The implications of stone tool types for the presence of language during the Lower and Middle Palaeolithic." In *The Human Revolution*, Paul Mellars and Chris Stringer (eds), 415–432. Edinburgh: Edinburgh University Press.
- Donald, Merlin. 1991. *Origins of the Modern Mind*. Cambridge (MA): Harvard University Press.
- Dubreuil, Benoit, in press. "The cognitive foundations of institutions." In *Beyond the brain: Embodied, situated and distributed cognition*, Benoit Hardy-Vallée and Nicolas Payette (eds.). Newcastle: Cambridge Scholars Publishing.
- Duday, Henri, 1995. "Anthropologie «de terrain»." In *La Mort passé, présent, conditionnel*, GVEP (ed), 33–58. La Roche-sur-Yon: GVEP.
- Duday, Henri and. Masset, Claude (eds). 1987. *Anthropologie physique et Archéologie. Méthodes d'études des sépultures*. Paris: CNRS.
- Dunbar, Robin I. M. 1996. *Grooming, gossip and the evolution of language*. London: Faber and Faber.
- Falk, Dean. 1987. "Brain lateralization in primates and its evolution in Hominids." *Yearbook of Physical Anthropology* 30: 107–125.
- Feuillet, Lionel, Dufour, Henry and Pelletier, Jean. 2007. « Brain of a white-collar worker." *The Lancet* 370: 262 – 262.
- Formicola, Vincenzo, Pettitt Paul B. and Del Lucchese, Angiolo. 2004. "A direct AMS radiocarbon date on the Barma Grande 6 Upper Paleolithic Skeleton." *Current Anthropology* 45 (1): 115–117.
- Forster, Peter. 2004. "Ice Ages and the mitochondrial DNA chronology of human dispersals: a review." *Philosophical Transactions of the Royal Society London Series B* 359: 255–64.
- Gambier, Dominique. 2000. "Aurignacian Children and Mortuary practice in Western Europe." *Anthropologie* 38 (1): 5–21.
- Gambier, Dominique. 1997. "Modern Human at the Beginning of the Upper Paleolithic in France: Range of Anthropological data and Perspectives." In *Conceptual Issues in Modern Human Origins research*, Clark Geoffrey A. and Catherine M. Willermet (eds), 117–131. New York: Aldine de Gruyter.
- Gambier, Dominique, Houet Francis and Tillier, Anne-Marie. 1990. « Dents de Font de Gaume (Châtelperronien et Aurignacien) et de La Ferrassie (Aurignacien ancien) en Dordogne. » *Paléo* 2: 143–152.
- Gardin, Jean-Claude, 1990. "The structure of archaeological theories." In *Mathematics and Information Science in Archaeology*, [Studies in Modern Archaeology 3], Albertus Voorrips (ed), 7–28. Bonn: Holos.
- Gargett, Robert H. 1989. "Grave shortcomings: the evidence for Neandertal burial." *Current Anthropology* 30: 157–190.
- Gargett, Robert H. 1999. "Middle Palaeolithic burial is not a dead issue: the view from Qafzeh, Sainte-Césaire, Kebara, Amud, and Dederiyeh." *Journal of Human Evolution* 37: 27–40.
- Garrod, Dorothy and Bate, Dorothea M. 1937. *The stone-age of Mount-Carmel. Vol. 1. Excavations at the Wadi-el-Mughara*. Oxford: Clarendon Press.
- Godfrey-Smith, Dorothy I. and Ilani, Shimon. 2004. "Past thermal history of goethite and hematite fragments from Qafzeh Cave deduced from thermal activation characteristics of the 110°C TL peak of enclosed quartz grains." *Revue d'Archéométrie* 28: 185–190.

- Goren-Inbar, Naama. 1986. "A figurine from the acheulian site of Berekhat Ram." *Journal of the Israel Prehistoric Society* 19: 7–12.
- Gravina, Brad, Mellars Paul A. and Bronk Ramsey, Christopher. 2005. "Radiocarbon dating of interstratified Neanderthal and early modern human occupations at the Chatelperronian type-site." *Nature* 438: 51–56.
- Grimaud-Hervé, Dominique. 1997. *L'évolution de l'encéphale chez Homo erectus et Homo sapiens*. Paris: CNRS.
- Grün, Rainer and Stringer, Chris B. 2000. "Tabun revisited: revised ESR chronology and new ESR and U-series analyses of dental material from Tabun C1." *Journal of Human Evolution* 39 (6): 601–612.
- Grün, Rainer, Stringer, Chris, McDermott, Frank, Nathan, Roger, Porat, Naomi, Robertson, Steve, Taylor Lois, Mortimer Graham, Eggins, Stephen and McCulloch, Malcolm. 2005. "U-series and ESR analyses of bones and teeth relating to the human burials from Skhul." *Journal of Human Evolution* 49: 316–334.
- Hahn, Joachim and Münzel, Susanne. 1995. "Knochenflöten aus dem Aurignacien des Geißenklösterle bei Blaubeuren, Alb-Donau-Kreis." *Fundberichte aus Baden-Württemberg* 20: 1–12.
- Heim, Jean-Louis. 1976. *Les Hommes fossiles de la Ferrassie*. [Archives de l'Institut de Paléontologie humaine, mémoire 35] Tome 1. Paris: Masson.
- Hein, Wulf. 1998. "Zur Rekonstruktion und Funktion jungpaläolithischer Knochenflöten." *Musica Instrumentalis* 1: 120–128.
- Henry-Gambier, Dominique. 2002. "Les fossiles de Cro-Magnon (Les-Eyzies- de-Tayac, Dordogne): Nouvelles données sur leur position chronologique et leur attribution culturelle." *Bulletin et Mémoires de la Société d'Anthropologie de Paris* 14 (1–2): 89–112.
- Henry-Gambier, Dominique, Maureille, Bruno and White, Randall. 2004. «Vestiges humains des niveaux de l'Aurignacien ancien du site de Brassempouy (Landes).» *Bulletins et Mémoires de la Société d'Anthropologie de Paris* 16: 49–87.
- Henshilwood, Christopher S. 2006. Stratigraphic integrity of the Middle Stone Age levels at Blombos Cave. In *From Tools to Symbols. From Early Hominids to Modern Humans*, Lucinda Backwell and d'Errico Francesco (eds). 441–458. Johannesburg, Witwatersrand University Press.
- Henshilwood, Christopher S. 2007. Fully symbolic sapiens behaviour: Innovation in the Middle Stone Age at Blombos Cave, South Africa. In *Rethinking the Human Revolution: New Behavioural and Biological Perspectives on the Origins and Dispersal of Modern Humans*, Chris Stringer & Paul Mellars (eds.) McDonald Institute Research Monograph series, 123–132, Cambridge, University of Cambridge Press.
- Henshilwood, Christopher S. 2008. *Holocene prehistory of the southern Cape, South Africa: excavations at Blombos Cave and the Blombosfontein Nature Reserve*. BAR S1860, Cambridge: Cambridge Monographs in African Archaeology 75: 1–171.
- Henshilwood, Christopher S. 2008. "Winds of change: palaeoenvironments, material culture, and human behavior in the late Pleistocene (~77ka -48ka ago) in the Western Cape Province, South Africa." *South African Archaeological Society Goodwin Series* 10: 1–17.
- Henshilwood, Christopher S. 2009. The origins of symbolism, spirituality & shamans: exploring Middle Stone Age material culture in South Africa. In *Becoming human: innovation in prehistoric material and spiritual cultures*, Colin Renfrew & Ian Morley (eds.) 29–49. Cambridge, Cambridge University Press.

- Henshilwood, Christopher S. 2009. The "Upper Palaeolithic" of southern Africa: the Still Bay and Howiesons Poort techno-traditions. In *African Genesis: Perspectives on hominid evolution*, Sally Reynolds & Andrew Gallagher (eds.). 38–50, Johannesburg: Wits University Press
- Henshilwood, Christopher S. and Marean, Curtis W. 2003. "The origin of modern human behaviour: A review and critique of models and test implications." *Current Anthropology* 44: 627–651.
- Henshilwood, Christopher S., d'Errico, Francesco, Marean, Curtis, Milo, Richard and Yates, Royden. 2002. "An early bone tool industry from the Middle Stone Age at Blombos Cave, South Africa: implications for the origins of modern human behaviour, symbolism and language." *Journal of Human Evolution* 41: 631–678.
- Henshilwood, Christopher S., d'Errico, Francesco, Vanhaeren, Marian, van Niekerk, Karen. and Jacobs, Zenobia. 2004. "Middle Stone Age Shell Beads from South Africa." *Science* 304: 403.
- Henshilwood, Christopher S., d'Errico, Francesco and Watts, Ian. 2009. "Engraved ochres from the Middle Stone Age levels at Blombos Cave, South Africa." *Journal of Human Evolution* 57(1): 27–47.
- Henshilwood, Christopher S., d'Errico, Francesco, Yates, Royden, Jacobs, Zenobia, Tribolo, Chantale, Duller, Geoff A.T., Mercier, Norbert, Sealy, Judith C., Valladas, Hélène, Watts, Ian and Wintle, Ann G. 2002. "Emergence of modern human behavior: Middle Stone Age engravings from South Africa." *Science* 295: 1278–1280.
- Henshilwood, Christopher S. and Dubreuil, Benoit, 2009. "Reading the artifacts: gleaned language skills from the Middle Stone Age in southern Africa." In *The Cradle of Human Language*, eds. Rudolf Botha and Chris Knight. 41–60. Oxford: Oxford University Press.
- Henshilwood, Christopher S., Sealy, Judith C., Yates, Royden J., Cruz-Urbe, Kathryn, Goldberg, Paul, Grine, Frederick E., Klein, Richard G., Poggenpoel, Cedric, van Niekerk, Karen L. and Watts, Ian. 2001. "Blombos Cave, southern Cape, South Africa: Preliminary report on the 1992 – 1999 excavations of the Middle Stone Age levels." *Journal of Archaeological Science* 28 (5): 421–448.
- Holloway, Ralph L. 1985. "The poor brain of *Homo sapiens neanderthalensis*: see what you please." In *Ancestors: the hard evidence*, Eric Delson (ed), 319–324. New York: Alan R. Liss.
- Holloway, Ralph L. 1995. "Toward a synthetic theory of human brain evolution." In *The origin of the Human Brain*, Jean-Pierre Changeux et Jean Chavaillon (eds), 42–60. Oxford: Clarendon Press.
- Hovers, Erella and Belfer-Cohen, Anna. 2005. «"Now You See it, Now You Don't" – Modern Human Behavior in the Middle Paleolithic.» In *Transitions before the Transition: Evolution and Stability in the Middle Paleolithic and Middle Stone Age*, Erella Hovers and Steve Kuhn (eds), 295–304. New York: Springer.
- Hovers, Erella, Ilani, Shimon, Bar-Yosef, Ofer and Vandermeersch, Bernard. 2003. "An Early Case of Color Symbolism. Ochre Use by Modern Humans in Qafzeh Cave." *Current Anthropology* 44 (4): 492–522.
- Hovers, Erella, Kimbel William H. and Rak Yoel. 2000. "The Amud 7 skeleton – still a burial: response to Gargett." *Journal of Human Evolution* 39 (2): 253–260.
- Hovers, Erella, Vandermeersch, Bernard. and Bar-Yosef, Ofer. 1997. "A Middle Palaeolithic engraved artefact from Qafzeh Cave, Israel." *Rock Art Research* 14: 79–87.
- Hublin, Jean-Jacques. 2005. "La langue des premiers Hommes." In *Aux origines des langues et du langage*, Jean-Marie Hombert (ed), 102–117. Paris: Fayard.
- Hublin, Jean-Jacques, Spoor, Fred, Braun, Marc, Zonneveld, Frans and Condemi, Silvana. 1996. "A late Neanderthal associated with Upper Palaeolithic artefacts." *Nature* 381: 224–226.

- Ingman, Max, Kaessmann, Henrik, Paabo, Svante and Gyllensten, Ulf. 2000. "Mitochondrial genome variation and the origin of modern humans." *Nature* 408: 708–713.
- Jackendoff, Ray and Pinker, Steven. 2005. "The nature of the language faculty and its implications for evolution of language. Reply to Fitch, Hauser, & Chomsky." *Cognition* 97: 211–225.
- Jackendoff, Ray. 1999. "Possible stages in the evolution of the language capacity." *Trends Cognitive Sciences* 3: 272–279.
- Jacobs, Zenobia. 2004. *Development of luminescence dating techniques for dating Middle Stone Age sites in South Africa*. PhD thesis. Aberystwyth: University of Wales.
- Jacobs, Zenobia, Duller, Goeff A.T., Wintle, Ann G. and Henshilwood, Christopher S. 2007. "Extending the chronology of deposits at Blombos Cave, South Africa, back to 140 ka using optical dating of single and multiple grains of quartz." *Journal of Human Evolution* 51 (3): 255–273.
- Jacobs, Z., Wintle, Ann G. and Duller, Goeff A.T. 2003. "Optical dating of dune sand from Blombos Cave, South Africa: I – multiple grain data." *Journal of Human Evolution* 44: 599–612.
- Jacobs, Zenobia, Roberts, Richard G., Galbraith, Rex F., Deacon, Hilary J., Grün, Rainer, Mackay, Alex, Mitchell, Peter, Vogelsang, Ralf, Wadley, Lyn. 2008. "Ages for the Middle Stone Age of Southern Africa: Implications for Human Behaviour and Dispersal." *Science* 322: 733–735.
- Jakobson, Roman. 1980. *Brain & Language. Cerebral Hemispheres and Linguistic Structure in Mutual Light*. [Interdisciplinary Series, Vol.4]. New York: New York University Slavic Papers.
- Jehs, Dieter. 2004. "Gauging Neanderthal cognition: problems and perspectives." *Ghent Archaeological Studies* 1: 85–99.
- Jouzel, Jean, Masson-Delmotte, Valerie, Cattani, Olivier, Dreyfus, Gabrielle, Falourd, Sonia, Hoffmann, G., Minster, B., Nouet, J., Barnola, Jean-Marc, Chappellaz, Jérôme A., Fischer, Hubertus, Gallet, J.C., Johnsen, Sigfus J., Leuenberger, M. C., Loulergue, Laetitia, Luethi, D., Oerter, Hans, Parrenin, Frédéric, Raisbeck, Grant, Raynaud, Dominique, Sclit, Adrian, Schwander, Jakob, Selmo, Enrico, Souchez, Roland, Spahni, Renato, Stauffer, Bernhard, Steffensen, Jorgen P., Stenni, Barbara, Stocker, Thomas F., Tison, Jean-Louis, Werner, M., Wolff, Eric W. 2007. "Orbital and Millennial Antarctic Climate Variability over the Past 800,000 Years." *Science* 317 (5839): 793–796.
- Jungers, William L., Pokempner, Amy A., Kay Richard F. and Cartmill, Matt. 2003. "Hypoglossal Canal Size in Living Hominoids and the Evolution of Human Speech." *Human Biology* 75: 473–484.
- Kay, Richard F., Cartmill, Matt and Balow, Michelle. 1998. "The hypoglossal canal and the origin of human vocal behavior." *Proceedings of the National Academy of Science* 95: 5417–5419.
- Klein, Richard G. 1999. *The human career*. 2nd edition. Chicago (IL): Chicago University Press.
- Klein, Richard G. 2000. "Archaeology and the evolution of human behaviour." *Evolutionary Anthropology* 9: 17–36.
- Klein, Richard G. 2003. "Whither the Neanderthals?" *Science* 229: 1525–1527.
- Klein, Richard G. and Edgar, Blake. 2003. *The Dawn of Human Culture*. New York: John Wiley & Sons.
- Knight, Chris. 2008. "Language co-evolved with the rule of law." *Mind & Society* 7: 109–128.
- Knight, Chris, Power, Camila and Watts, Ian. 1995. "The Human Symbolic Revolution: A Darwinian Account." *Cambridge Archaeological Journal* 5: 75–114.
- Krause, Johannes, Lalueza-Fox, Carles, Orlando, Ludovic, Enard, Wolfgang, Green, Richard E., Burbano, Hernan A., Hublin, Jean-Jacques, Hänni, Catherine, Fortea, Javier, de la Rasilla, Marco, Bertranpetit, Jean, Rosas, Antonio and Pääbo, Svante. 2007. "The Derived FOXP2

- Variant of Modern Humans Was Shared with Neandertals." *Current Biology* 17: 1908–1912.
- Kuhn Steve L., Stiner, Mary C., Reese, David S. and Güleş, Erksin. 2001. "Ornaments of the earliest Upper Paleolithic: new insights from the Levant." *Proceeding of the National Academy of Science USA* 98 (13): 7641–7646.
- Kunej, Drago and Turk, Ivan. 2008. "New Perspectives on the Beginning of Music: Archeological and Musicological Analysis of a Middle Paleolithic Bone «Flute»." In *The Origins of Music*, Nils L. Wallin, Björn Merker and Steven Brown (eds), 234–268. Cambridge (MA): MIT Press.
- Lahr, Martha M., Foley, Robert A. 1998. "Towards a Theory of Modern Human Origins: Geography, Demography, and Diversity in Recent Human Evolution." *Yearbook of Physical Anthropology* 41: 137–176.
- Laitman, Jeffrey T., Reidenberg, Joy S., Gannon, Patrick J., Johansson, B., Landahl, Karen and LeMay, Marjorie. 1975. "The language capability of Neanderthal man." *American Journal of Physical Anthropology* 42: 9–14.
- Lalueza-Fox, Carles, Römpler Holger H., Caramelli, David, Stäubert, Claudia, Catalano, Giulio, Hughes, David, Rohland, Nadin, Pilli, Elena, Longo, Laura, Condemi, Silvana, de la Rasilla, Marco, Fortea, Javier, Rosas, Antonio, Stoneking, Mark, Schöneberg, Torsten, Bertranpetit, Jaume and Hofreiter, Michael. 2007. "A Melanocortin 1 Receptor Allele Suggests Varying Pigmentation Among Neanderthals." *Science* 318: 1453–1455.
- Lambert, Fabrice, Delmonte, Barbara, Petit, Jean-Robert, Bigler, Matthias, Kaufmann, Patrick R., Hutterli, Manuel A., Stocker, Thomas F., Ruth, Urs, Steffensen, Jorgen P. and Maggi, Valter. 2008. "Dust-climate couplings over the past 800,000 years from the EPICA Dome C ice core." *Nature* 452 (7187): 616–619.
- Lawson, Graeme and d'Errico Francesco. 2002. "Microscopic, experimental and theoretical reassessment of Upper Palaeolithic bird-bone pipes from Isturitz, France: ergonomics of design, systems of notation and the origins of musical traditions." In *Studien zur Musikarchäologie III. Archäologie früher Klangerzeugung und Tonordnung*, Ellen Hickmann, Anne D. Kilmer and Ricardo Eichmann (eds), 119–142 [Orient-Archäologie Band 10]. Rahden Westfalen: Verlag Marie Leidorf.
- Lawson, Graeme, 2004. "Music, Intentionality and Tradition: identifying purpose, and continuity of purpose, in the music-archaeological record." In *Studien zur Musikarchäologie III. Archäologie früher Klangerzeugung und Tonordnung*, Ellen Hickmann, Anne D. Kilmer and Ricardo Eichmann (eds), 61–97 [Orient-Archäologie Band 10]. Rahden Westfalen: Verlag Marie Leidorf.
- Lawson, Graeme, in press. In *Ian Cross and Bennett Zon (eds) Durham Music Evolution conference 2006*.
- Leakey, Mary. 1971. *Olduvai Gorge Excavations in Beds I and II, 1960–1963*. Cambridge: Cambridge University Press.
- Leavesley, Matthew G. 2007. "A shark-tooth ornament from Pleistocene Sahul." *Antiquity* 81: 308–315.
- Lerdahl, Fred and Jackendoff, Ray. 1983. *A Generative Theory of Tonal Music*. Cambridge (MA): MIT Press.
- Leroi-Gourhan, Arlette and Leroi-Gourhan, André. 1965. "Chronologie des grottes d'Arcy-sur-Cure." *Gallia Préhistoire* 7: 1–64.
- Leroi-Gourhan, André. 1964. *Les religions de la préhistoire*. Paris: PUF.
- Lewis-Williams, David. 2002. *The Mind in the Cave: Consciousness and the Origins of Art*. London: Thames & Hudson.

- Li, Jun Z., Absher, Devin M., Tang, Hua, Southwick, Audrey M., Casto, Amanda M., Ramachandran, Sohini, Cann, Howard M., Barsh Gregory S., Feldman, Marcus, Cavalli-Sforza Luigi L. and Myers, Richard M. 2008. "Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation." *Science* 319 (5866): 1100 – 1104.
- Lieberman, Philip. 1975. *On the origins of language: an introduction to the evolution of human speech*. New York (NY): Macmillan.
- Lieberman, Philip. 1989. "The origins of some aspects of human language and cognition." In *The human revolution: behavioural and biological perspectives on the origins of modern humans*, Paul A. Mellars and Chris B. Stringer (eds), 391–414. Princeton (NJ): Princeton University Press.
- Lieberman, Philip. 2001. "On the neural bases of spoken language." In *In the mind's eye: multidisciplinary approaches to the evolution of human cognition*, April Nowell (ed), 172–186 [Monographs 13]. Ann Arbor (MI): International Monographs in Prehistory.
- Lieberman, Philip. 1984. *The Biology and Evolution of Language*. Cambridge MA: Harvard University Press
- Lieberman, Philip. 2006. *Toward an evolutionary biology of language*. Cambridge (MA): Harvard University Press.
- Lieberman, Philip. 2007. "Current views on Neanderthal speech capabilities: A reply to Boë *et al.* (2002). Letter to the editor." *Journal of Phonetics* 35: 552–563.
- Mackay, Alex and Welz, Aara. 2008. "Engraved ochre from a Middle Stone Age context at Klein Kliphuis in the Western Cape of South Africa." *Journal of Archaeological Science* 35: 1521–1532.
- Maclarnon, Ann and Hewitt, Gwen. 2004. "Increased breathing control: another factor in the evolution of human language." *Evolutionary Anthropology* 13: 181–197.
- Marean, Curtis W., Bar-Matthews, Miryam, Bernatchez, J., Fisher Erich, Goldberg, Paul, Herries, Andy I.R., Jacobs, Zenobia, Jerardino, Antonieta, Karkanas, Panagiotis, Minichillo, Tom, Nilssen, Peter J., Thompson, Erin, Watts, Ian and Williams, Hope M. 2007. "Early human use of marine resources and pigment in South Africa during the Middle Pleistocene." *Nature* 449 (7164): 905–908.
- Marshack, Alexander. 1976. "Some Implications of the Paleolithic Symbolic Evidence for the Origin of Language." *Current Anthropology* 17: 274–282.
- Marshack, Alexander. 1981. "On Paleolithic Ochre and the Early Uses of Color and Symbol." *Current Anthropology* 22: 188–191.
- Marshack, Alexander. 1996. "A Middle Palaeolithic symbolic composition from the Golan Heights: the earliest known depictive image." *Current Anthropology* 37: 357–65.
- Martínez, Ignacio, Arsuaga, Juan Luis, Quam, Rolf, Carrettero, José Miguel, Gracia, Ana and Rodríguez, Laura. 2008. "Human hyoid bones from the middle Pleistocene site of the Sima de los Huesos (Sierra de Atapuerca, Spain)." *Journal of Human Evolution* 54 (1): 118–124.
- Martínez, Ignacio, Rosa, Manuel, Arsuaga, Juan Luis, Jarabo, Pilar, Quam, Rolf, Lorenzo, Carlos, Gracia, Ana, Carrettero, José Miguel, Bermúdez de Castro José Maria and Carbonell, Eudald. 2004. "Auditory capacities in Middle Pleistocene humans from the Sierra de Atapuerca in Spain." *Proceedings of the National Academy of Sciences* 101: 9976–9981.
- Maureille, Bruno and Van Peer, Philip. 1998. "Une donnée peu connue sur la sépulture du premier adulte de La Ferrassie (Savignac-de-Miremont, Dordogne)." *Paléo* 10: 291–302.
- Maureille, Bruno. 2004. *Les origines de la culture: les premières sépultures*. Paris: Le Pommier.
- Maureille, Bruno and Vandermeersch, Bernard. 2007. «Les sépultures néandertaliennes.» In *Les Néandertaliens, biologie et cultures*, [Documents préhistoriques 23], Bernard Vandermeersch and Bruno Maureille, (eds), 311–322. Paris: C.T.H.S.

- Maureille, Bruno and Tillier, Anne-Marie. 2008. « Répartition géographique et chronologique des sépultures néandertaliennes. » In *Première humanité, gestes funéraires des Néandertaliens*, Bernard Vandermeersch, Jean-Jacques Cleyet-Merle, Jacques Jaubert, Bruno Maureille and Alain Turq, (eds), 66–74. Paris: Réunion des Musées Nationaux.
- McBrearty, Sally and Brooks Alison S. 2000. “The revolution that wasn’t: a new interpretation of the origin of modern human behaviour.” *Journal of Human Evolution* 39: 453–563.
- McBrearty, Sally, 2001. “The Middle Pleistocene of East Africa.” In *Human roots: Africa and Asia in the Middle Pleistocene*, Lawrence Barham and Kate Robson-Brown, (eds), 81–92. Bristol: Western Academic and Specialist Press.
- McBurney, Charles B. M. 1967. *The Haua Fteah (Cyrenaica) and the Stone Age of the south-east Mediterranean*. Cambridge: Cambridge University Press.
- McCown, Theodore D. 1937. “Mugharet es-Skhul, description and excavations.” In *The stone age of the Mount Carmel. Excavations at the Wady el-Mughara*, Dorothy A. E. Garrod and Dorothea M. A. Bate (eds), 91–113. Oxford: Clarendon Press.
- McCown, Theodore D. and Keith, Arthur. 1939. *The Stone Age of Mount Carmel II: The Fossil Human Remains from the Levallois-Mousterian*. Oxford: Clarendon Press.
- McDougall, Ian, Brown, Francis H. and Fleagle, John G. 2005. “Stratigraphic placement and age of modern humans from Kibish, Ethiopia.” *Nature* 433: 733–736.
- McPherron, Shennan, Soressi, Marie and Dibble, Harold. 2001. « Deux nouveaux projets de recherche à Pech de l’Azé (Dordogne, France). » *Préhistoire du Sud-Ouest* 8: 11–30.
- Mellars, Paul. 1996a. *The Neandertal legacy: an archaeological perspective from western Europe*. Princeton (NJ): Princeton University Press.
- Mellars, Paul. 1996b. “Symbolism, language, and the Neanderthal mind.” In *Modelling the early human mind*, Paul Mellars and Kathleen Gibson (eds) 15–32. Cambridge: McDonald Institute.
- Mellars, Paul. 1998. “Neandertals, Modern Humans and the archaeological evidence for language.” In *The origin and diversification of language*, Nina G. Jablonski and Leslie C. Aiello (eds), 89–115 [Mémoires 24]. San Francisco (CA): California Academy of Sciences.
- Mellars, Paul. 2005. “The impossible coincidence: a single-species model for the origins of modern human behavior in Europe.” *Evololutionary Anthropology* 14: 12–27.
- Mellars, Paul and Gibson, Kathleen (eds). 1996. *Modelling the early human mind*. Cambridge: McDonald Institute.
- Mellars, Paul, Gravina, Brad and Bronk Ramsey, Christopher. 2007. “Confirmation of Neanderthal/modern human interstratification at the Chatelperronian type-site.” *Proceedings of the National Academy of Sciences USA* 104(9): 3657–3662.
- Mellars, Paul and Stringer, Chris, (eds). 1989. *The Human Revolution: Behavioral and Biological Perspectives on the Origins of Modern Humans*. Edinburgh: Edinburgh University Press.
- Mellars, Paul, Boyle, Katty, Bar-Yosef, Ofer and Stringer, Chris (eds). 2007. *Rethinking the Human Revolution*. Cambridge: McDonald Institute Monographs, 275–286.
- Mercier, Norbert, Valladas, Hélène, Froget, Laurence, Joron, Jean-Louis and Ronen, Avraham. 2000. « Datation par la thermoluminescence de la base du gisement paléolithique de Tabun (Mont Carmel, Israël). » *Comptes Rendus de l’Académie des Sciences* 330: 731–738.
- Mithen, Steve J. 1996. *The prehistory of the mind: a search for the origins of art, religion and science*. London: Thames & Hudson.
- Mithen, Steve J. 1999. “Symbolism and the supernatural.” In *The Evolution of Culture*, Robin Dunbar, Chris Knight and Power, Camila (eds), 147–169. Edinburgh: Rutgers University Press.
- Morel, Jean. 1974. « Nouvelles datations absolues de formations littorales et de gisements préhistoriques de l’Est algérien. » *Bulletin de la Société préhistorique française* 71: 103–105.

- Morley, Iain R. 2006. « Archaeological implications of intentionality in hunter-gatherer music.” In *Archaeoacoustics*, Chris Scarre and Graeme Lawson, (eds) 95–105. Cambridge: McDonald Institute Monographs.
- Morse, Kate. 1993. “Shell beads from Mandu Mandu Creek rock-shelter, Cape Range peninsula, Western Australia, dated before 30,000 BP.” *Antiquity* 67: 877–883.
- Nettle, David. 1998. “Explaining Global Patterns of Language Diversity.” *Journal of Anthropological Archaeology* 17: 354–374.
- Noble, William and Davidson, Ian. 1996. *Human evolution, language and mind: a psychological and archaeological inquiry*. Cambridge: Cambridge University Press.
- Nowell, April and d’Errico, Francesco. 2007. “The Art of Taphonomy and the Taphonomy of Art: Layer IV, Molodova I, Ukraine.” *Journal of Archaeological Method and Theory* 14: 1–26.
- O’Connor, Sue. 1995. “Carpenter’s Gap 1 rockshelter: 40000 years of aboriginal occupation in the Napier Ranges, Kimberley, W.A.” *Australian Archaeology* 40: 58–59.
- Ogawa, T., Kamiya, T., Sakai, S. and Hosokawa, H. 1970. Some observations on the endocranial cast of the Amud man, in *The Amud Man*, eds. H. Suzuki & F. Takai. Tokyo: University of Tokyo Press, 411–424.
- Palma di Cesnola, Arturo. 1993. *Il Paleolitico superiore in Italia*. Firenze: Garlatti e Razzai.
- Patel, Aniruddh D. 2008. *Music, Language, and the Brain*. Oxford: Oxford University Press.
- Pavlov, Pavel, Svendsen, John Inge and Indrelid, Svein. 2001. “Human presence in the European Arctic nearly 40,000 years ago.” *Nature* 413: 64–67.
- Pearson, Michael P. 1999. *The Archaeology of Death and Burial*. Phoenix Mill: Sutton.
- Pettitt, Paul B., Richards, Martin P., Formicola, Vincenzo and Maggi, Roberto. 2003. “The Gravettian burial known as ‘The Prince’ (‘Il Principe’): new evidence for his age and diet.” *Antiquity* 77: 15–19.
- Pettitt, Paul B. and Bader Nicolai O. 2000. “Direct AMS Radiocarbon dates for the Sungir mid Upper Palaeolithic burials.” *Antiquity* 74: 269–270.
- Pettitt, Paul. 2002. “The Neanderthal dead: exploring mortuary variability in Middle Palaeolithic Eurasia.” *Before Farming* 1 (44): 1–19.
- Peyrony, Denis. 1934. «La Ferrassie, Moustérien, Périgordien, Aurignacien.» *Préhistoire* 3: 1–92.
- Pinker, Steven. 1997. *How the Mind Works*. London: Allen Lane.
- Pinker, Steven. 2000. *The Language Instinct: How the Mind Creates Language*. New York (NY): Harper Perennial Modern Classics.
- Piperno, Marcello and Giacobini, Giacomo. 1992. “A taphonomic study of the paleosurface of Guattari Cave.” In *Proceedings of the International Symposium The fossil man of Monte Circeo: fifty years of studies on the neandertals in Latium*, Amilcare Bietti and Giorgio Manzi (eds), 143–161. Roma: Quaternaria Nova.
- Ponce de Leon, Marcia S., Golovanova, Lubov, Doronichev, Vladimir, Romanova, Galina, Akazawa, Takeru, Kondo, Osamu, Ishida Hajime and Zollikofer Christopher P. E. 2008. “Neanderthal brain size at birth provides insights into the evolution of human life history.” *Proceedings of the National Academy of Sciences* 105: 13764–13768.
- Powell, Adam, Shennan, Stephen and Thomas, Mark. 2009. “Late Pleistocene Demography and the Appearance of Modern Human Behavior.” *Science* 324 (5932): 1298–1301.
- Power, Camila and Aiello, Leslie. 1997. “Female Proto-symbolic Strategies.” In *Women in Human Evolution*, Lori D. Hager (ed), 153–171. London, New York: Routledge.
- Rak, Yoel, Kimbel, William H. and Hovers, Erella. 1994. “A. Neanderthal infant from Amud Cave, Israel.” *Journal of Human Evolution* 26: 313–324.

- Renfrew, Collin A. 1996. "The sapient behaviour paradox: how to test for potential?" In *Modelling the early human mind*, Paul Mellars and Kathleen Gibson (eds), 11–14. Cambridge: McDonald Institute.
- Renfrew, Collin and Zubrow, Ezra B. W. 1994. *The Ancient Mind: Elements of Cognitive Archaeology*. Cambridge, UK: Cambridge University Press.
- Richards, Michael P., Pettitt, Paul B., Stiner, Mary C. and Trinkaus, E. 2001. "Stable isotope evidence for increasing dietary breadth in the European mid-Upper Paleolithic." *Proceedings of the National Academy of Science USA* 97: 7663–7666.
- Rigaud, Jean-Philippe, Texier, Pierre-Jean, Parkington, John and Poggenpoel, Cedric. 2006. «Le mobilier Stillbay et Howiesons Poort de l'abri Diepkloof: la chronologie du Middle Stone Age sud-africain et ses implications.» *Comptes Rendus Palevol* 5: 839–849.
- Rigaud, Solange, d'Errico, Francesco, Vanhaeren, Marian and Neumann, Christian 2009. "Critical reassessment of putative Acheulean *Porosphaera globularis* beads." *Journal of Archaeological Science* 36 (1): 25–34.
- Rizzi, L., 2009. "The adaptive approach to grammar." In *Biological Foundations and Origin of Syntax*, Derek Bickerton and Eörs Szathmáry (eds) [Strüngmann Forum Reports 3]. Cambridge (MA): MIT Press.
- Rodriguez, Laura, Cabo, Luis L. and Egocheaga, José Enrique. 2003. "Breve nota sobre el hioides Neandertalense de Sidron (Pilona, Asturias)." In: Maria Pilar Aluja, Assumpcio Malgosa, Ramon N. Nogués (eds), 484–493 [Antropología y Diversidad, 1]. Barcelona: Edicions Bellaterra.
- Rockman, Marcy and Steele, James (eds). 2003. *Colonization of unfamiliar landscapes: the archaeology of adaptation*. Oxford: Routledge.
- Roux, Valentine and Bril, Blandine (eds). 2005. *Stone knapping: the necessary conditions for a uniquely hominin behaviour*. Cambridge (UK): McDonald Institute for Archaeological Research.
- Roux, Valentine. 2007. "Ethnoarchaeology: A Non Historical Science of Reference Necessary for Interpreting the Past." *Journal of Archaeological Method and Theory* 14 (2): 153–178.
- Ruhlen, Merritt. 1994. *The Origin of Language: Tracing the Evolution of the Mother Tongue*. New York (NY): John Wiley & Sons.
- Rumbaugh, Duane M. and Washburn David A. 2003. *Intelligence of Apes and Other Rational Beings*. New Haven (CT): Yale University Press.
- Šajnerová-Dušková, Andrea, Fridrich, Jan and Fridrichová-Sýkorová, Ivana. 2009. "Pitted and grinding stones from Middle Palaeolithic settlements in Bohemia: a functional study." In *Non-flint Raw Material Use in Prehistory: Old Prejudices and New Directions*, Farina Sternke, Lotte Eigeland and aurent-Jacques Costa (eds), [Proceedings of the XV Congress of the U.I.S.P.P.] Oxford: Archaeopress.
- Schepartz, Lynne A. 1993. "Language and modern human origins." *Yearbook Physical Anthropology* 36: 91–126.
- Searle, John R. 1995. *The Construction of Social Reality*. New York: Free Press.
- Semendeferi, Katerina and Damasio, Hanna. 2000. "The brain and its anatomical subdivision using magnetic resonance imaging." *Journal of Human Evolution* 38: 317–332.
- Semendeferi, Katerina. 2001. "Before or after the split? Hominid neural specialisations." In *In the Mind's Eye*, April Nowell (ed), 107–120 [International Monographs in Prehistory]. Michigan: Ann Arbor.
- Shennan, Stephen. 2001. "Demography and Cultural Innovation: a Model and its Implications for the Emergence of Modern Human Culture." *Cambridge Archaeological Journal* 11: 5–16.

- Sladek, Vladimir, Trinkaus, Eric, Hillson, Simon W. and Holliday, Trenton W. 2000. *The people of the Pavlovian* [The Dolní Vestonice Studies 5]. Brno: Institute of Archaeology, Academy of Sciences of the Czech Republic.
- Soressi, Marie and d'Errico, Francesco. 2007. "Pigments, gravures, parures: les comportements symboliques controversés des Néandertaliens." In *Les Néandertaliens. Biologie et cultures*, Bernard Vandermeersch and Bruno Maureille, (eds), 297–309. Paris: Éditions du CTHS.
- Soressi, Marie, Jones, Heather L., Rink, W. Jack, Maureille, Bruno and Tillier, Anne-Marie. 2007. «ESR, Uranium Series and AMS 14C dating of Pech de l'Azé 1". *Journal of Human Evolution*, 52: 455–466.
- Steguweit, L. 1999. "Intentionelle Schnittmarken auf Tierknochen von Bilzingsleben: Neue la-
sermikroskopische Untersuchungen." *Praehistoria Thuringica* 3: 64–79.
- Stout, Dietrich, Toth, Nicholas, Schick, Kathy, Stout, Julie and Hutchins Gary. 2000. "Stone Tool-Making and Brain Activation: Position Emission Tomography (PET) Studies." *Journal of Archaeological Science* 27 (12): 1215–1223.
- Stout, Dietrich and Chaminad, Thierry. 2007. "The evolutionary neuroscience of tool making." *Neuropsychologia* 45 (5): 1091–1100.
- Stringer, Chris. 2002. "Modern human origins – progress and prospects." *Philosophical transactions of the Royal Society, London (B)* 357: 563–579.
- Stringer, Chris B. and Gamble, Clive. 1993. *In search of the Neandertals*. London: Thames and Hudson.
- Svoboda, Jiri A., Van Der Plicht, Johannes and Vítězslav Kuzela, V. 2002. "Upper Palaeolithic and Mesolithic human fossils from Moravia and Bohemia (Czech Republic): some new C14 dates." *Antiquity* 76: 9577–962.
- Számádó, Szabolcs, Bishop, Dorothy, Deacon, Terrence, d'Errico, Francesco, Fischer, Julia, Hurford, James, Okanoya, Kazuo, Szathmáry, Eörs and White, Stephany. 2009. "What are the possible biological and genetic foundations for syntactic phenomena?" In *Biological Foundations and Origin of Syntax*, Derek Bickerton and Eörs Szathmáry (eds) [Strüngmann Forum Reports 3].
- Taborin, Yvette. 2003. «La mer et les premiers hommes modernes.» In *Echanges et diffusion dans la préhistoire méditerranéenne*, Bernard, Vandermeersch (ed), 113–122. Paris: CTHS.
- Taçon, Paul. 2004. "Ochre, clay, stone and art. The symbolic importance of minerals as life force among aboriginal peoples of Northern and Central Australia." In *Soil, Stones and Symbols: Cultural Perceptions of the Mineral World*, Nicole Boivin and Mary Ann Owoc (eds), 30–42. London: UCL Press.
- Tattersal, Ian. 1999. *Becoming Human: Evolution and Human Uniqueness*. Harcourt: Brace.
- Tattersall, Ian and Schwartz, Jeffrey H. 1999. "Hominids and hybrids: The place of Neanderthals in human evolution." *Proceedings of the National Academy of Sciences USA* 96: 7117–7119.
- Templeton, Alan, 2005. "Haplotype trees and modern human origins." *Yearbook of Physical Anthropology* 48: 33–59.
- Thomas, Downing A. 1995. *Music and the Origins of Language: Theories from the French Enlightenment*. Cambridge (UK): Cambridge University Press.
- Tillier, Anne-Marie, Arensburg, Baruch, Vandermeersch, Bernard and Rak Y. 1991. «L'apport de Kébara à la paléthnologie funéraire des Néandertaliens du Proche-Orient.» In *Le Squelette moustérien de Kébara* 2, Ofer Bar-Yosef and Bernard Vandermeersch (eds), 89–95. Paris: CNRS.
- Tillier, Anne-Marie and Arensburg, Baruch. 2000. "The speech capability of Neandertals: an obsolete question." *Prehistoria* 1: 55–62.

- Tillier, Anne-Marie. 2005. "The Tabun C1 Skeleton: A Levantine Neanderthal?" *Journal of the Israel Prehistoric Society* 35: 439–450.
- Tobias, Philip V. 1987. "The brain of *Homo habilis*: A new level of organisation in cerebral evolution." *Journal of Human Evolution* 16 (7/8): 741–761.
- Tobias, Philip V. 2005. "Tools and brains: which came first?" In *From Tools to Symbols. From Early Hominids to Modern Humans*, Francesco d'Errico and Lucinda Backwell. (eds.), 82–102. Johannesburg: Wits University Press.
- Tomasello, Michael, Carpenter, Malinda, Call, Josep, Behne Tanya and Moll, Henrike. 2005. "Understanding and sharing intentions: The origins of cultural cognition." *Behavioral and Brain Sciences* 28: 675–735.
- Trinkaus, Eric, Moldovan, Oana, Milota, Stefan, Bilgar, Adrian, Sarcina, Laurentiu, Athreya, Sheela, Bailey, Shara E., Rodrigo, Ricardo, Mircea, Gherase, Higham, Tom, Bronk Ramsey Christopher and Van der Plicht, Johannes. 2003. "An Early Modern Human from Pestera cu Oase, Romania." *Proceedings of the National Academy of Science USA* 100, 20: 11231–11236.
- Trinkaus, Eric, 2005. "Early modern humans." *Annual Review of Anthropology* 34: 207–230.
- Trinkaus, Eric and Zilhão, João. 2002. "Phylogenetic implications." In *Portrait of the artist as a child: the Gravettian human skeleton from the Abrigo do Lagar Velho and its archeological context*, João Zilhão and Eric Trinkaus (eds), 497–518 [Trabalhos do Arqueologia 22]. Lisboa: Instituto Português de Arqueologia, Gobierno de Cantabria/Parlamento de Cantabria.
- Tschauner, Hartmut, 1996. "Middle-range theory, behavioral archaeology, and post-empiricist philosophy of science in archaeology." *Journal of Archaeological Method and Theory* 3: 3–30.
- Turk, Ivan (ed). 1997. *Mousterian "Bone Flute"? and Other Finds from Divje Babe I Cave Site in Slovenia*. Ljubljana: Institut za Arhaeologijo.
- Van Peer, Philip and Vermeersch Pierre M. 2000. "The Nubian Complex and the dispersal of modern humans in North Africa." In *Recent Research into the Stone Age of Northeastern Africa*, Lech Krzyzaniak, Karla Kroeper and Michal Kobusiewicz (eds), 47–60 [Studies in African Archaeology 7]. Poznan: Poznan Archaeological Museum.
- Van Peer, Philip, Fullagar, R., Stokes, Steven, Bailey, R.M., Moeyersons, Jan, Steenhoudt, Frans, Geerts, A., Vanderbeken, T., De Dapper, M. and Geus, F. 2003. "The early to Middle Stone Age transition and the emergence of modern human behavior at site 8-B-11, Sai Island, Sudan." *Journal of Human Evolution* 45 (2): 187–193.
- Van Peer, Philip, Rots, Veerle and Vroomans, Jeanne-Marie. 2004. "A story of colourful diggers and grinders: the Sangoan and Lupemban at site 8-B-11, Sai Island, Northern Sudan." *Before Farming* 3: 1–28.
- Vandermeersch, Bernard. 1993. Appendix, in *Context of a late Neandertal*, eds. François, Lévêque, Anna Mary Backer and Michel Guilbaud (eds), 129–131 [Monographs in World Archaeology 16]. Madison: Prehistory Press.
- Vandermeersch, Bernard. 1969. «Les nouveaux squelettes moustériens découverts à Qafzeh (Israël) et leur signification.» *Comptes Rendus de l'Académie des Sciences de Paris*, 268, série D: 2562–2565.
- Vandermeersch, Bernard. 1981. *Les Hommes Fossiles de Qafzeh (Israel)*. Paris: Centre National de la Recherche Scientifique.
- Vandermeersch, Bernard, Cleyet-Merle, Jean-Jacques, Jaubert, Jacques, Maureille, Bruno and Turc, Alain (eds). 2008. *Première humanité, gestes funéraires des Néandertaliens*. Paris: Réunion des Musées Nationaux.

- Vanhaeren, Marian. 2005. "Speaking with beads: The evolutionary significance of bead making and use." In *From Tools to Symbols. From Early Hominids to Modern Humans*, Francesco d'Errico and Lucinda Backwell (eds), 525–535. Johannesburg: Wits University Press.
- Vanhaeren, Marian and d'Errico, Francesco. 2006. "Aurignacian ethnolinguistic-geography of Europe revealed by personal ornaments." *Journal of Archaeological Science* 33: 1105–1128.
- Vanhaeren, Marian and d'Errico, Francesco, Stringer, Chris, James, Sarah L., Todd, Jon and Mienis, Henk K. 2006. "Middle Palaeolithic Shell Beads in Israel and Algeria." *Science* 312: 1785–1788.
- Vermeersch, Pierre M., Gijssels, Gilbert and Paulissen, Etienne. 1984. "Discovery of the Nazlet Khater man, Upper Egypt." *Journal of Human Evolution* 13: 281–286.
- Vermeersch, Pierre M., Paulissen, Etienne, Stokes, Steven, Charlier, C., Van Peer, Philip, Stringer, Chris B. and Lislay, W. 1998. "A Middle Palaeolithic burial of a modern human at Taramsa Hill, Egypt." *Antiquity* 72: 475–484.
- Villa, Paola, Delagnes, Anne and Wadley, Lyn. 2005. "A late Middle Stone Age artifact assemblage from Sibudu (KwaZulu-Natal): Comparisons with the European Middle Paleolithic." *Journal of Archaeological Science* 32: 399–422.
- Villa, Paola, Soressi, Marie, Henshilwood, Chris and Mourre, Vincent. 2009. "The Still Bay points of Blombos Cave (South Africa)." *Journal of Archaeological Science* 36 (2): 441–460.
- Wadley, Lyn. 2001. "What is cultural modernity? A general view and a South African perspective from Rose Cottage." *Cambridge Archaeological Journal* 11 (2): 201–221.
- Wadley, Lyn. 2003. "How some archaeologist recognize culturally modern behaviour." *South African Journal of Science* 99: 247–250.
- Wadley, Lyn. 2006. "Partners in grime: results of multidisciplinary archaeology at Sibudu cave." *Southern African Humanities* 18 (1): 315–341.
- Walter, Philippe. 2003. «Caractérisation des traces rouges et noires sur les coquillages perforés de Qafzeh.» In *Echanges et Diffusion dans la Préhistoire méditerranéenne*, Bernard Vandermeersch (ed), 122. Paris: CTHS.
- Watts, Ian. 1999. "The origin of symbolic culture." In *The evolution of culture*, Robin Dunbar, Chris Knight and Camila Power (eds), 113–146. Edinburgh: Edinburgh University Press.
- Watts, Ian. 2002. "Ochre in the Middle Stone Age of southern Africa: ritualized display or hide preservative?" *South African Archaeological Bulletin* 57: 1–14.
- Watts, Ian. 2009. "Red ochre, body painting and language: interpreting the Blombos ochre." In *The cradle of language*, Rudolf Botha and Chris Knight (eds), 62–92. Oxford: Oxford University Press.
- Watts, Ian. in press. "The pigments from Pinnacle Point Cave 13B, Western Cape, South Africa." *Journal of Human Evolution*.
- White, Tim D. 1986. "Cut marks on the Bodo cranium: a case of prehistoric defleshing." *American Journal of Physical Anthropology* 69 (4): 503–509.
- White, Tim D. 2000. "Cutmarks on the Bodo cranium: A case of prehistoric defleshing." *American Journal of Physical Anthropology* 69: 503–550.
- White, Tim and Nicolas Toth. 1991. "The Question of Ritual Cannibalism at Grotta Guattari." *Current Anthropology* 32 (2): 118–138.
- White, Tim D., Asfaw, Berhane, DeGusta, David, Gilbert, Henry, Richards, Gary D., Suwa, Gen and Clark, Howell F. 2003. "Pleistocene *Homo sapiens* from Middle Awash, Ethiopia." *Nature*, 423: 742–47.
- White, Randal. 2001. "Personal ornaments from the Grotte du Renne at Arcy-sur-Cure." *Athena Review* 2: 41–46.

- Whiten, Andrew. 2005. "The second inheritance system of chimpanzees and humans." *Nature* 437: 52–55.
- Wolpoff, Milford. 1999. *Paleoanthropology*. New York: McGraw-Hill.
- Wreschner, Ernst. 1981. "More on Palaeolithic Ochre." *Current Anthropology* 22: 705–706.
- Wurz, Sarah, 2000. *The Middle Stone Age at Klasies River, South Africa*. Ph.D Dissertation: University of Stellenbosch.
- Wynn, Thomas and Coolidge, Fred L. 2004. "The expert Neanderthal mind." *Journal of Human Evolution* 46: 467–487.
- Zilhão, João and d'Errico, Francesco. 1999. "The chronology and taphonomy of the earliest Aurignacian and its implications for the understanding of Neandertal extinction." *Journal of World Prehistory* 13 (1): 1–68.
- Zilhão, João and d'Errico, Francesco. 2003a. "An Aurignacian «garden of Eden» in southern Germany? An alternative interpretation of the Geissenklösterle and a critique of the Kulturpumpe model." *Paléo* 15: 69–86.
- Zilhão, João and d'Errico, Francesco. 2003b. "The chronology of the Aurignacian and Transitional technocomplexes. Where do we stand?" In *The Chronology of the Aurignacian and of the Transitional Technocomplexes. Dating, Stratigraphies, Cultural Implications* [Trabalhos de Arqueologia 33], João Zilhão and Francesco d'Errico (eds), 313–349. Lisboa: IPA.
- Zilhão, João and Trinkaus, Eric (eds). 2003. *Portrait of the Artist as a Child. The Gravettian Human Skeleton from the Abrigo do Lagar Velho and its Archeological Context* [Trabalhos de Arqueologia 22]. Lisboa: IPA.
- Zilhão, João, d'Errico, Francesco, Bordes, Jean-Guillaume, Lenoble, Arnaud, Texier, Jean-Pierre and Rigaud, Jean-Philippe. 2008. "Grotte des Fées (Châtelperron): history of research, stratigraphy, dating and archaeology of the Châtelperronian type-site." *Paleoanthropology* 2008: 1–42.
- Zilhão, João, d'Errico, Francesco, Bordes, Jean-Guillaume, Lenoble, Arnaud, Texier, Jean-Pierre and Rigaud, Jean-Philippe. 2006. "Analysis of Aurignacian interstratification at the Châtelperronian-type site and implications for the behavioral modernity of Neandertals." *Proceedings of the National Academy of Sciences* 103: 12643–12648.
- Zubrov, Ezra and Blake, Elizabeth. 2006. "The Origin of Music and Rhythm." In *Archaeoacoustics*, Chris Scarre and Graeme Lawson (eds), 117–126. Cambridge (UK): McDonald Institute Monographs.

Early diffusion of domestic bovids in Europe

An indicator for human contacts, exchanges and migrations?

Anne Tresset¹, Ruth Bollongino^{1,2}, Ceiridwen J. Edwards³,
Sandrine Hughes⁴ and Jean-Denis Vigne¹

Cattle, sheep and goat were domesticated in the Near-East during the 9th millennium BC. From there, sheep and goat, which had no wild ancestors in Europe, were introduced to this continent at the beginning of the 7th millennium B.C. and diffused following two main flows: a southern route along the northern coastline of the Mediterranean, and a northern route across central Europe following the Danubian corridor. Possible scenarios of migration have been complicated to investigate regarding cattle, as the species had a possible wild ancestor in Europe: the local aurochs, whose disappearance only occurred at the end of the 17th century A.D. and whose remains are hardly distinguishable from those of the early domestic forms on the basis of classical osteometry. A tight cooperation between Archaeozoology and Genetics has provided, in the frame of several publicly funded projects (among which the OMLL scheme), substantial new data allowing refinement of historical scenarios to a degree never achieved thus far. We were able to demonstrate that local aurochs did not contribute, or contributed to a very limited extent, to the constitution of European domestic cattle herds, whose origin can be clearly traced back to the Near East. Thus, from this point of view, domestic cattle biogeographical history is very similar to sheep and goat, and their appearance in Europe probably owes more to farming pioneers than to local hunter-gatherers. Analyses of goat aDNA revealed the preservation of an important genetic diversity very far from the diffusion centre.

1. CNRS, Archéozoologie, Archéobotanique: Sociétés, Pratiques, Environnements. UMR 7209 du CNRS/USM 303 MNHN, 55 Rue Buffon, BP 56, 75005 Paris, France.

2. Palaeogenetics Group, Institut für Anthropologie, Johannes Gutenberg-Universität Mainz, Colonel-Kleinmann-Weg 2, 55128 Mainz, Germany.

3. Research Laboratory for Archaeology, Oxford University, South Parks Road, Oxford OX1 3QY, UK.

4. Paleogenetics and Molecular Evolution: Institut de Génomique Fonctionnelle de Lyon, Université de Lyon 1, CNRS, INRA, Lyon 1; Ecole Normale Supérieure de Lyon, France.

This is suggestive of the persistence of gene flow between domestic herds across the dispersion area along the different diffusion routes, which prevented the occurrence of severe bottleneck effects. This diversity also indicates that the existence of contacts between farming groups encompassed very large areas. It is very interesting to note that recent works published on domestication and diffusion of pig in Neolithic Europe have proposed very different scenarios. This highlights the specificity of domestic bovids as tracers of human contacts, exchanges and displacements during the Neolithisation of Europe.

Anthropic dissemination of exotic species as a tracer of human movements and contacts.

During the Holocene, the human introduction of new species to Europe began with the dissemination of the Neolithic culture, and culminated in modern times with the European discovery of New Worlds and the multiplication of intercontinental travels (Pascal *et al.* 2005, 2006). In the latter case, as evidenced by written documents, the circulation of animals and plants mirrors clearly the history of human movements, intercultural contacts and exchanges. This seems to have been especially true for domestic species, both plants and animals, and for taxa that could be utilised as pets or in pharmacopeia, as these had to be introduced with attached lore, implying some form of cultural exchange and transfer of technical knowledge. It is more than likely that comparable processes – even if not identical, the historical situation being very different – took place with the introduction of domestic plants and animals during Neolithic times to Europe. Thus tracing domestic animal movements accompanying the Neolithic dissemination, and documenting their rhythm, should shed light on the nature of the neolithisation phenomenon itself, as a process of diffusion of ideas, cultures and people.

The origin of domestic bovids in Europe: A not so simple question.

The origin of domestic bovids in Europe (*i.e.* sheep, *Ovis aries*; goat, *Capra hircus*; cattle, *Bos taurus*) is not an easy question to tackle for several reasons. It is essentially a twofold problem: firstly, the identification of wild ancestor populations at the origin of domestication (their distribution in time and space and their biological traits), and secondly, the description of characteristics (rhythm, main vectors, etc.) of the dissemination process of domesticated animals across Europe, which we know at least partly contributed to the appearance of the first domestic herds on this continent. Domestic bovids cannot be treated as a whole in Europe:

sheep and goat had no wild ancestors in this part of the world (Poplin 1979; Uerpmann 1979) and, therefore, their appearance on this continent is necessarily consecutive to their domestication. Cattle (as well as pig), conversely, had potential ancestors in Europe; the distribution of aurochs during the first part of the Holocene included most of this territory (Van Vuure 2005). Thus European aurochs could have contributed to the constitution of European Neolithic herds, by independent local domestication or by hybridisation with cattle imported from the Near East. Another difficult question is determining the method of diffusion of these animals into Europe: we know that domesticates were moved by the farmers themselves during the Neolithic colonisation process of Europe, but we do not know if the late Mesolithic people contributed to this diffusion by the adoption of husbandry practices (which is not unlikely, considering modern well-documented cases of hunter communities taking up husbandry in a context of colonisation by farmers; review in Digard 1990), or more punctually by acquiring domesticates (dead or alive) by exchange or theft, thus introducing them to new territories.

Building up refined and testable scenarios typically requires an interdisciplinary approach, as in most research on the origins of farming, but perhaps with additional insight: the question of ancestry is more a 'naturalistic' one, requiring investigation of life sciences and palaeobiology; the diffusion process is more an 'historical' question that can only be documented by archaeological observation. However, each aspect – naturalistic and historical – has implications to the other in the process of animal domestication and diffusion. They are intertwined and so should be treated together. This is why interdisciplinary efforts, joining experts from different academic fields (archaeology, archaeozoology and molecular genetics) is essential.

At the beginning of the story: The Near East.

Domestication of wild ungulates can be detected by the main visible biological consequence, that of size decrease of the skeleton, as well as by age and sex composition of the individuals that have been consumed by humans, the refuse of which has accumulated at archaeological sites (see e.g. Vigne *et al.* 2005). Using both of these criteria, Peters *et al.* (2005) evidenced the earliest domestication of sheep and goat in the world on the Southern slopes of the Eastern Taurus Mountains, in South-Eastern Anatolia, Turkey, and dated it ca. 8500 cal. BC (Figure 1). Both the wild ancestor of sheep (the oriental mouflon, *Ovis orientalis*), and goat (the bezoar goat, *Capra aegagrus*), were living in this area, together with the ancestors of domestic cattle, the aurochs *Bos primigenius*, and several other ancestors of domesticated plants and animals (see ref. in Harris 1996). Sheep and goat were

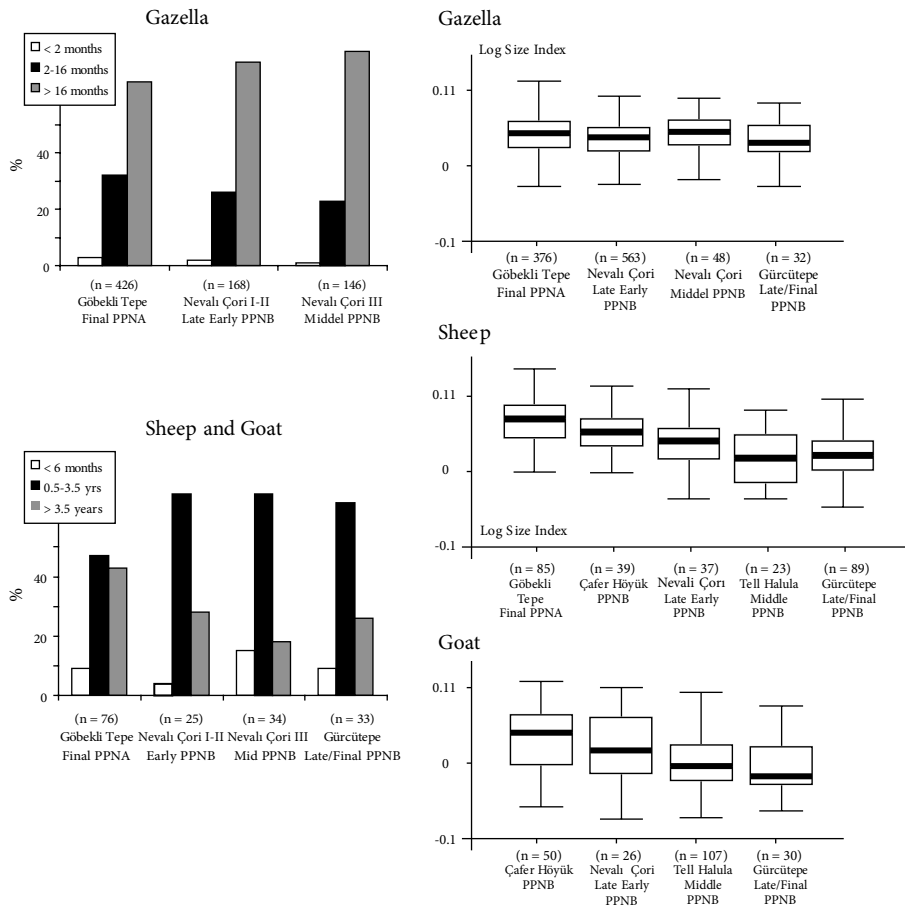


Figure 1. Comparison of the evolution of the culling profiles (left) and of the general body size (right; estimated by log-size index) of gazella (*Gazella subgutturosa*), sheep and goat from the early to recent PPNB (ca. 9000 to 7000 BC) in the upper Euphrates valley (South-eastern Anatolia, Turkey), modified from Peters et al. (2005). For gazelle, neither the culling profiles nor the size changed during these two millennia. Conversely, the ratio between sub-adults and adults in the slaughtering profile strongly, and for the long term, increased for sheep/goat (which could not be distinguished based on the tooth remains by the authors) in the middle of the ninth millennium. At the same time, the body size of both sheep and goat begun to decrease significantly; this did not result from any climatic or landscape change, since the body size was not modified at the same time. Concomitant and sustainable modifications of slaughtering profiles and size decrease are good evidence for the domestication of sheep and goat ca. 8500 BC in this area. This is its earliest known attestation

domesticated by groups of early Pre-Pottery Neolithic (PPN) villagers. These village societies were strong and well organised, with very high technological skills, e.g. flint and obsidian knapping, and had been sedentary for four millennia and were practicing cultivation of cereal and leguminous for several centuries (Tanno & Willcox 2005; Vigne 2008). They also had community buildings in each village, and specialised ceremonial sites common to several villages, such as Göbeckli, characterised by a very rich artistic repertoire dominated by animal representations (Helmer *et al.* 2004; Peters & Schmidt 2004). The relative homogeneity of the early PPN material culture from the high (Turkey) to the middle (Syria) Euphrates Valley confirms broad exchanges and mobility between villages.

Cattle domestication has long been considered to have occurred later as no size decrease is detectable in bovine bones dating to the ninth millennium. Only the transportation of cattle to Cyprus, through a 70 to 80 km boat trip, as early as

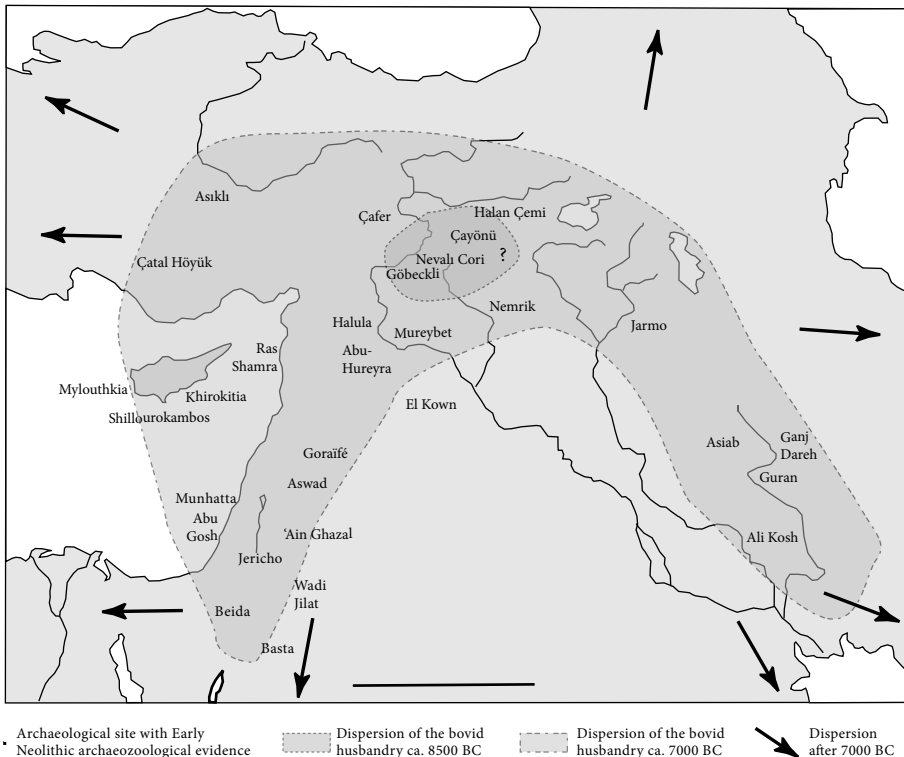


Figure 2. Location of main Near Eastern early Neolithic sites that have provided archaeozoological evidence, approximate delimitation of the areas of dispersion of the bovid domestication ca. 8500 BC (Early PPNB) and ca. 7000 BC (final Late PPNB), and main streams of diffusion after 7000 BC

8300 cal. BC (Vigne *et al.* 2000, 2003) suggests earlier management. However, applying more refined morphological analyses to the bovid bone remains of Djâde (Syria), Helmer *et al.* (2005) detected a significant decrease of the sexual dimorphism ca. 8500–8300 cal. BC, a biological modification which is known to occur in several vertebrate species as soon as they live in captivity (Arbuckle 2005). Consequently, nowadays, although most archaeozoologists consider that sheep, goat and cattle (and probably pig, as well) were domesticated in the Upper Euphrates and Tigris valleys during the middle of the ninth millennium by early PPNB villagers, there is also agreement that bovid domestication took place at different places (Figure 2). The most convincing evidence is the domestication of goat in the Zagros (Iran) in the early eighth millennium (Zeder 2005) and of cattle in the lower Indus valley (Pakistan) during the seventh millennium (Meadow 1984). These two independent domestications stem from the local wild populations of bezoar goat and zebu aurochs, respectively. The Zagros and Indus valley are both located several thousands kilometres away from the Eastern Taurus and the contexts of domestication are characterised there by later and very different human cultures. However, long distance interactions between late Mesolithic people of these regions and early Neolithic cultures of the Taurus cannot be completely excluded. In other words, we do not know yet if these domestications were totally independent from the ones in the Taurus (and one from each other), or if they were stimulated by eastward transfers of ideas and technical knowledge. This question remains open, as a crucial issue for understanding long distance interactions between people as early as the ninth to seventh millennia in the Near and Middle East.

In the Zagros, Zeder (2005) highlighted goat domestication based on demographic criteria alone, and did not detect any size decrease during the first millennium of goat management. Size decrease is also very slight and hardly detectable in the early cattle introduced to Cyprus (Vigne *et al.* 2003), and was not evidenced in the sites from South-Eastern Anatolia (Peters *et al.* 2005). This leaves open another question: can archaeology detect very early stages of animal management, *i.e.* did animal domestication begin much earlier than the ninth millennium with no visible biological modification of animals? Several archaeologists have, despite no univocal evidence, already proposed ‘pre-domestications’ or attempts at ungulate managing during the twelfth to ninth millennia (e.g. Legge 1972; Redding 2005).

Recent results obtained for goat in the Middle East suggest that population genetics may provide unexpected evidence about very early animal managements before ‘true’ domestication. Genetic analysis of 473 samples collected over the whole distribution range of the modern bezoar goats (Naderi *et al.* 2008) shows a Late Glacial or early Holocene expansion of the wild lineages of the Zagros area that subsequently gave rise to domestic goats. This suggests a widespread management involving hundreds of wild goat populations before the earliest domestication

sites detected by archaeology. Some of these managed goat populations were obviously transferred out of their source area at that time or shortly after. This scenario differs from the traditional model that imagines domestication occurring within restricted centres, bringing about strong genetic bottleneck effects. Although it is not certain if this goat model can be transposed to the other ungulates, it does lend a very interesting renewed view of exchanges of skills and animals between human communities in this area, before the 'true' Neolithic period. Traditionally, a domestic population is considered to be characterised by lower genetic diversity than wild populations, depending on how many individuals were initially captured and how isolated they were kept from their wild relatives. The example of goat shows that factors like genetic isolation from the wild population, growth and expansion of the domestic population and possible interbreeding with domesticates from other regions could have caused an increasing differentiation between the domestic animals and their wild source population in the course of time.

Only a large amount of ancient DNA data would allow us to describe these processes in detail, but, with regard to the Near East, unfortunately the local climate does not favour the preservation of DNA (e.g. Edwards *et al.* 2004). In hot climates, the bones are subject to severe thermal stress (Smith *et al.* 2003; Bollongino & Vigne 2007). Furthermore, in most regions of the Near East, the amount of precipitation is enough to keep the soil humid almost all year round and this enhances bone degradation (see Nielsen-Marsh & Hedges 2000a, b for a summary of main degradation factors). Additionally, the damaged structure of bone leaves the samples vulnerable to contamination and special caution during sample treatment is crucial (Bollongino *et al.* 2008a).

First diffusions, first acclimatisation and new technical skills

For goat, translocations of managed or domestic animals before the mid-ninth millennium seem to have been restricted to the distribution range of the wild ancestor (Naderi *et al.* 2008). Conversely, from 8500–8300 cal. BC onwards, archaeozoological data attest to long distance transfers (Figure 2). Goats, cattle and pigs were imported to Cyprus before 8300 BC, with sheep and the Mesopotamian fallow deer (*Dama mesopotamica*) appeared on the island at the very beginning of the eighth millennium (Vigne *et al.* 2005; Vigne in press). Domestic sheep and goats are attested at the same time in the Middle Euphrates valley (Halula, Abu Hureyra) where there were no wild sheep and possibly no wild goat present (Peters *et al.* 1999). All four domestic ungulates have even recently been dated to ca. 8300 BC several hundreds kilometres southwards, at Tell Aswad, Damascus (Helmer & Gourichon in press). The dry Mediterranean environment of Cyprus, and moreover

the semi-arid plain of Damascus, should have entailed very different conditions for early domestic bovids with reference to ones living alongside their wild ancestors on the slopes of the Taurus mountains. This suggests that PPNB Neolithic people should have had to find new management techniques for acclimatising them in other areas (Vigne 2008). For our purpose, the most important implication is that this transfer to new areas required transmission of greater packages of knowledge between human groups that, though pertaining more or less to the same PPNB cultural complex, are clearly identified as different from each other on the basis of their material cultures.

These remarks are still more relevant for the subsequent steps of the diffusion process. In short, at the end of the PPNB (ca. 7000 BC), the area of bovid husbandry was extended from Central Anatolia to the Negev Desert, and from the Mediterranean shores to the Western half of the Iranian Plateau (Figure 2). From this vast area, animal domestication spread to North-Eastern Africa, to the Arabic area, to the Indus valley, to Central Asia and to Europe, at the turn of the seventh to sixth millennia. It is clear that husbandry practices had to be adapted to each of the different local conditions and human culture traditions would have been very diverse across such large areas. Ancient DNA data are still too scarce to trace the routes of diffusion within the Near and Middle East and beyond. Most of the time, even archaeological data are not numerous enough to bring detailed information about these complex processes of local re-composition of the initial Neolithic package. However, they should not deeply differ from the ones in Europe, where data are much more numerous.

Europe as a new world

Small stock

Sheep and goat were introduced to South-Eastern Europe at the very beginning of the seventh millennium BC. As non-native species in Europe, they can be considered as good paradigms of the advance of farming during the seventh to fourth millennia on this continent or, at least, as proof of Neolithic influence when found in too small proportions to establish the presence of husbandry practices. Thus, sheep and goat occurrences contribute to updating chronology and routes of diffusion of early domesticates and/or farming across Europe (Figure 3, Tresset & Vigne 2007). Regarding sheep and goat, new work has concerned Romania, where several early sites were recently sampled, the earliest dates being ca. 6000 BC (Balasescu & Tresset *et al.* work in progress) or earlier, and are probably the oldest obtained for domestic animals north of the Danube.

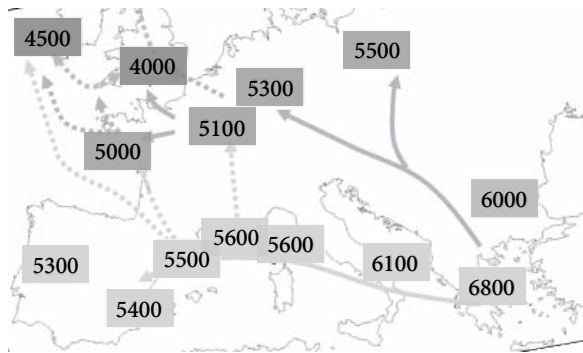


Figure 3. Main flows of domesticate diffusion during the Neolithisation of Europe. All dates are in years cal. BC. Modified from Tresset & Vigne (2007)

Sheep and goat are often considered together, indistinctly: although good methodological advances have recently been made in osteology to distinguish bone and tooth remains from the two species on a morphological basis (Halstead *et al.* 2002; Helmer 2000; Balasse & Ambrose 2005), the main part of small stock bone assemblages remains in the category ‘indeterminate Caprini’. This situation could mask dissimilarities between the two species within the diffusion process, in terms of dissemination dynamics or importance in husbandry strategies, including the adaptation to new territories. Despite these limitations, recent investigations on the Northern coastline of the Mediterranean, where sheep overwhelmingly outnumbered goats, highlight the transportation of domestic animals during the earliest phase of the Neolithic. In this area, the earliest phase of the Neolithic is represented by the *Impressa* cultural complex, traditionally known to have been restricted to Italy and dated to 6200–5500 BC. In Southern France and Spain, archaeologists have long considered that the earliest manifestation of the Neolithic was represented by the Cardial-Epicardial cultures, dated to 5400–4800 BC. Starting from the 1990s, they began to realise that a pioneer *Impressa* wave had also reached Southern France leaving only a few archaeological sites (13 at the present time) along the coast and lower Rhône valley. The first monographic study of one of these sites, Pont de Roques-Hautes (Portiragnes, Hérault, France), dated ca. 5600 BC, sheds light on this ‘colonial’ process (Guilaine *et al.* 2007). Neolithic pioneer villagers brought with them their entire cultural package, including flint knapping, ware decoration, domestic cereals and well controlled husbandry dominated by specialised sheep breeding. The refined analysis of potteries and lithic industries demonstrate that this human group came from a rather restricted geographic area in Western-South/Central Italy. Morphometric investigations on the bones of sheep showed the same results, excluding the slender sheep of Northern

Italy. However, it seems that these pioneer coastal groups contributed very little to the constitution of the Cardial-Epicardial flocks, where the slightly robust *Impressa* sheep seem to have been totally replaced by the slender ones of Northern Italy. Even the economic system of the pioneer rapidly changed, turning to a non-specialised production based upon game hunting, and cattle and sheep husbandry.

This scenario illustrates well the massive local re-composition of the Neolithic package according not only to the local ecological conditions, but also overall to the local cultural traditions (Tresset & Vigne 2007). It also reveals a succession of waves of neolithisation, from different regions, and spanning through the Neolithic and later. It suggests long distance and frequent transportation of exotic matters, including cereals and sheep/goat, wide diffusion of knowledge, and frequent contacts between human populations. Conversely, the high level of local re-composition of the Neolithic package pleads for little migration of humans. These complex Neolithic systems of exchange appear to have been mainly based on gradual diffusions of materials and knowledge.

Small stock, among which sheep were probably the most abundant, were also an important component of the package diffused by early farmers along the continental ('Danubian') route of Neolithisation, even if often outnumbered by cattle. These species had to be adapted to climates differing strongly from their original environments. Moisture, temperature, but also daylight lengths, which depend on the latitude, have important impacts on bovid physiology, especially on the mating season of sheep and goat. The change to new environments under higher latitudes may have resulted in major dietary shifts and reproduction rescheduling. Recent work on these two aspects, based on serial analysis of stable isotopes within tooth crowns of sheep and cattle, suggests that this adaptation was completed by the turn of the third millennium BC in North-Western Europe (Balasse *et al.* 2006; Balasse & Tresset 2007), where farming was introduced relatively late (ca. 4000 cal. BC). The most spectacular Neolithic adaptation of sheep to high latitude and marine environment has been observed in animals from the site of Holm of Papa Westray (Orkney, ca. 3000 cal. BC), which were fed on seaweed during winter times, despite the fact that sheep are naturally not able to eat such food without severe health problems. This is the earliest record of a practice that is widespread nowadays in North-Western Europe, where it is also applied to other species (cattle, horse and pig).

Molecular approaches have yielded important information about the diffusion process of goats (that could be unambiguously identified as such, the difference with sheep being obvious from a genomic point of view). The mitochondrial (mtDNA) genetic diversity of present-day goats, and its spatial distribution, has been extensively studied, principally by sequencing the control region (nearly 2500 individuals analysed and more than 1500 haplotypes obtained; Naderi *et al.* 2007).

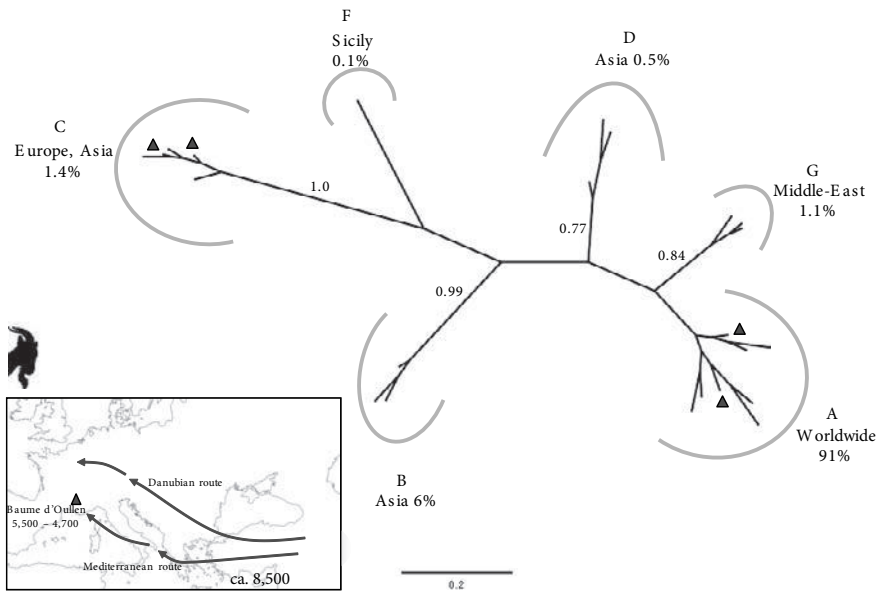


Figure 4. Present-day mitochondrial diversity of domestic goats and diversity observed in layers 6–5 of Baume d'Oullen, an early Neolithic archaeological deposit from Ardèche, France. 1. The phylogenetic tree was generated by Bayesian analyses (HKY-I-G model of evolution) on a reduced dataset of 23 sequences of 130 base pairs from the mt control region. 19 sequences have haplotypes used as references of the six maternal haplogroups (A, B, C, D, F, G) identified from present-day goats, and for which the current spatial distribution and the percentage of individuals are indicated (Naderi et al. 2008). The four remaining sequences (grey triangles) are those obtained from ancient domestic goat bones coming from layers 6–5 at Baume d'Oullen (Fernandez et al. 2006). Numbers above the branches of the tree are the posterior probabilities. 2. The map below indicates the two main routes of the initial advancement of the Neolithic culture into Europe from the Middle-East and the location of the Baume d'Oullen site (grey triangle). The dates are radio-carbon date-derived cal. BC

These studies have revealed that at least six major maternal haplogroups can now be described (Figure 4) and that the phylogeographical structuring is low compared to other livestock species. This weak structure could partly be explained by the worldwide distribution of one of the haplogroups (haplogroup A totals 91 per cent of all individuals sampled), but also indicates that goats were highly traded and their distribution influenced by human migration. The timing of these exchanges can be roughly estimated by present-day genetic studies but, here again, a more straightforward way is to use paleogenetic studies to have direct access to

genetic diversity through time. In this way, ancient goat bones of Baume d'Oullen, an early Neolithic archaeological site in Southern France, were analysed and a short fragment of the mitochondrial control region amplified for 19 samples (Fernandez *et al.* 2006). Two different maternal haplogroups (A and C) were present at the same time, in the same site, and far away from the initial domestication area in the Near East (goats have no wild ancestors in Europe, as we already emphasised). This observation suggests that a high gene flow existed at the beginning of the diffusion of domestic goats in Europe, which displaced the great diversity that had been generated by early goat management in the Near East (see above; Naderi *et al.* 2008).

In the present state of knowledge, there is no unquestionable evidence of sheep or goat in a Mesolithic context in Western Europe. Old claims at the sites of Gramari or Chateaufort-les-Martigues have been rejected a long time ago (Poplin *et al.* 1986), and those at Dourgne and Gazel, Southern France, or at Cova Fosca, Western Spain, have been questioned (Vigne & Helmer 1999; Vigne in Guilaine *et al.* 2007). The revision of the bone assemblages of the site of Beg an Dorchenn (Finistère, Brittany), dated to the mid- to late-sixth millennium, has shown that the sheep and/or goat remains that had been attributed to the late Mesolithic were very likely Iron Age intrusions (Tresset 2000, 2003).

Cattle

The existence of a potential progenitor for domestic cattle in Europe makes the question far more complex for bovines than for sheep and goat. The existence of local domestication centres in Europe and/or of hybridisation of the European aurochs with domestic cattle imported from the Near East has been much debated recently. The hypothesis of a local contribution of the aurochs to the European breeds has mostly relied on the large variation of bovine size recorded at early Neolithic sites, and the existence of 'intermediate' size groups situated between data interpreted as clearly referring to domestic cattle on the one hand, and to aurochs on the other. In this context, molecular approaches have been particularly helpful.

The first study of mtDNA variation on extant Eurasian cattle (Troy *et al.* 2001) revealed that the genetic diversity was highest within Near and Middle Eastern populations. The fact that European cattle consist of a subset of the West-Asian lineages, only at a far lower diversity, supports the archaeozoological findings of a Near Eastern centre of domestication. At the time of publication (2001), there were two major reasons why inferences concerning a Near Eastern origin of European cattle were treated with caution: (i) the data could not be easily extrapolated into the past, as the Neolithic diversity might have been overwritten by historic breeding practices; (ii) the European aurochs population had not been fully characterised, thus a European domestication centre could not be excluded. The

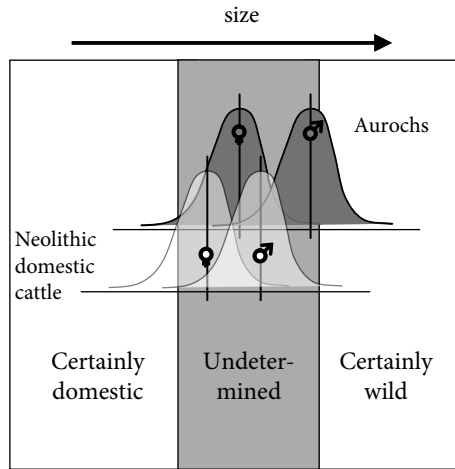


Figure 5. Schematic view of the general overlap between size ranges recorded for the domestic and wild forms of bovids at the beginning of the Neolithic

question of the origin of European cattle and a possible impact of European aurochs remained open. The fact that prehistoric domestic animals are generally smaller than wild ones (see above) encouraged numerous archaeozoological studies to address this question. However, the significant sexual dimorphism of bovids complicates the interpretation of osteometric data, as a medium-sized bone could either belong to a female aurochs or to a domestic bull, the body size variation of both forms overlapping widely (Figure 5). The presence of intermediate sized bones has been interpreted in two opposing ways. On the one hand, it was taken as a sign of a European domestication (or at least of crossbreeding between imported cattle and local aurochs) as medium sized animals were seen as a result of close relationship between the two forms, or at least a continuous influence of wild animals on domestic herds. On the other hand, opponents of this theory argued that the average body size of domestic animals was too small for a recent local domestication or broad hybridisation. In summation, neither modern genetic nor osteometric data are sufficient means to reveal the details of the population history of European cattle, pointing out the need for ancient DNA.

Ancient mitochondrial sequences of about 200 wild and domestic bovids are available (e.g. Edwards *et al.* 2004, 2007; Bollongino *et al.* 2006). The majority of the samples stem from Europe, spanning from the Palaeolithic to Bronze Age. The most striking finding is that European aurochs belong to a different population than domestic cattle (Figure 6, Edwards *et al.* 2007). The aurochs lineage cannot be found amongst today's European cattle, with very rare possible exceptions

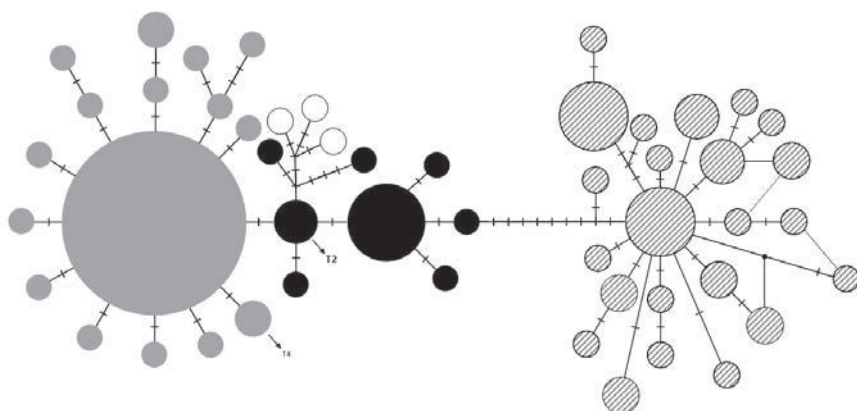


Figure 6. Median network showing 138 bovine haplotypes from European and Near Eastern Mesolithic, Neolithic and Bronze Age sites. The circles are scaled to the amount of individuals belonging to the respective lineage. The taurine lineages are shown in grey (T3), white (T1) and black (T), whereas the European aurochs lineages (P) are shown hatched. The European aurochs population reveals a great genetic distance to the taurine population which comprises domestic animals from Europe and the Near East

(see Achilli *et al.* 2008). The divergence time of European aurochs and cattle is of the order of 50,000 years (Ho *et al.* 2008), thus the two populations had already split prior to the onset of domestication. This implies that European aurochs were not the ancestors of domestic cattle. The few data that could be retrieved from Near Eastern samples belong to the taurine population of domestic cattle (Edwards *et al.* 2004, 2007). These results clearly corroborate the West Asian origin of European cattle. Figure 6 also reveals that the Neolithic diversity of cattle is close to modern times, suggesting that modern European cattle populations still reflect the prehistoric bottleneck that was caused by domestication and importation from the Near East. The fact that the lineage of the European aurochs is nearly completely missing amongst extant European cattle shows that female aurochs were not introduced to domestic herds on a common basis.

However, these findings need to be supplemented by more data from the Iberian and Italian peninsulas, where genetic diversity of aurochs is still under question (Edwards *et al.* 2007; Achilli *et al.* 2008). In addition, they are based on mitochondrial DNA, which is maternally inherited and reflects only the female part of the history. So far, no conclusions about a possible impact of aurochs bulls could be made. The analysis of Y-chromosomal patrilineages is hindered by a reduced amount of samples (only males have Y-chromosomes) and by a far worse preservation of nuclear DNA in archaeological samples. Götherström and colleagues (2005)

analysed Y-chromosomal loci of modern cattle and extinct aurochs and observed two different lineages, which were assumed to differentiate between European aurochs bulls and imported Near Eastern lineages. The distribution of these lineages amongst modern cattle suggested a major impact of local aurochs in Northern Europe. However, further testing of these markers in ancient samples revealed that the two lineages were both prevalent in European aurochs bulls (Svensson & Götherström 2008; Bollongino *et al.* 2008), and therefore the question of aurochs male introgression in Europe still remains unanswered on this basis.

Another hint at the existence of hybrids was the small size of some bovines from late Mesolithic sites located close to the limit of the contemporary Neolithic extension. This is the case at Rosenhof, Northern Germany, where a proportion of the bovine bones from the Ertebølle levels (fifth millennium cal. BC) fall at the lower limit of the size ranges known for local aurochs. The mtDNA analysis of nine of these bones yielded typically European aurochs sequences ('P' haplogroup; Scheu *et al.* 2007; Edwards *et al.* 2007). Thus, it could be imagined that the small size of these remains resulted from an introgression of imported cattle in the aurochs population. As demonstrated above, in the current state of knowledge, the analysis of Y-chromosomal patrilineages does not help to solve this question. However, analysis of the sex-linked Zinc Finger Gene of these bones allowed sex determination (Scheu *et al.* 2007). It turned out that all the 'small' remains were female, and thus it seemed likely that these were actually female aurochs, the size variation of which had been incompletely documented to date (data collected in Degerbøl & Fredskild 1970 for Denmark included very few females), rather than hybrids. Nevertheless, Geigl (2008) suggests a symmetrical situation (an introgression of wild genes into the domestic population) in Neolithic contexts from Northern France. Based on the analysis of amelogenin to identify sex of bovine remains, she found the existence of unusually large domestic cows (belonging to T haplogroup) and suggested that hybridisation between male aurochs and female domesticated cattle occurred at low frequency in Northern France during the Neolithic and Bronze Age. However, more data are needed to ascertain the frequency of this event, probably rare as this introgression seems to have left little if any evidence in modern domestic herds (see above).

Even if the hypotheses of local domestication and local hybridisation on a large scale can be reasonably ruled out, or at least minimised with the help of molecular genetics, this does not exclude a contribution of local Mesolithic populations to the process of introduction of domestic bovines to new territories, at least in the form of foodstuff. For example, there are claims of domestic cattle remains associated with Mesolithic industries on different sites in Brittany and Ireland (review in Tresset 2003). Regarding Brittany, Tresset (2000, 2003) has shown that on the site of Beg an Dorchenn, small bovines – attributed to the domestic form in

the literature – were probably Iron Age intrusions, as was the case on the same site for sheep and/or goat (see above). However, in the 1990s, two bovines, presumably domestic if we rely on horn morphology and general size, were discovered below the cairn of Er Grah at Locmariaquer (Morbihan), in a palaeosoil radiocarbon dated to the end of the sixth millennium (Tresset & Vigne 2006). This date locally corresponds to the end of the Mesolithic, but is also contemporary with Neolithic sites located a few hundreds kilometres to the south and east. Thus, there is a possibility that they were imported in Mesolithic territories as a result of contacts with Neolithic farmers. Among Irish sites, direct dating conducted by Woodman *et al.* (1997) confirmed the presence of cattle bones in association with late Mesolithic industries at the site of Ferriter's Cove, Kerry, in the mid-fifth millennium BC, predating the arrival of Neolithic cultures by a few centuries. The Irish situation offers some parallels with Brittany, as described above, but in this case, the transfer of animals from the continent requires skills in open sea sailing. As no evidence of such skills exists in the Mesolithic of North-Western Europe, Tresset (2003) suggested that continental Neolithic expeditions transporting cattle could have taken place only a few centuries before the Neolithic settlement of the island. Such expeditions would have left virtually no evidence if limited to the coast, because of subsequent changes to the coastline due to the Holocene sea-level rise, and the silting-up of estuaries and bays. Contemporary Mesolithic groups could have acquired domestic cattle by contact with these farming pioneers (via exchanges, gifts or thefts) or by hunting, after the feralisation of the newly introduced domestic bovines. Such a hypothesis is not as improbable as it sounds: feralisation of domestic animals introduced to new territories is no exception, and even seems to be a rule (review in Digard 1990), for example, in Corsica, the feralisation of domestic sheep, probably early within the Neolithic period, led to the emergence of the Corsican mouflon.

Conclusions

Despite their different situations in view of the distribution of their wild progenitors, sheep, goat and cattle have parallel Neolithic stories in Europe: all were introduced to this continent from the Near East with no (for sheep and goat) or extremely limited (for cattle) interaction with local wild bovid populations. Regarding cattle, such a definitive conclusion is new and could not have been reached without the help of palaeogenetics. The fact that all domestic bovids were introduced to Europe and constituted the basis of our numerous extant breeds is amazing in itself, but it is all the more so if we take into account the fact that these animals had to be acclimatised, sometimes to environments differing to the extreme from their source

regions. Recent data also show that this adaptation was complete by the end of the Neolithic expansion. Introducing, diffusing and adapting animals on a new continent required not only high zootechnical skills, but also the faculty to accumulate new observations, to adapt the common zootechnical knowledge to new conditions, and to be able to transmit it from one generation to the next, as well as between farming communities, as a cumulative knowledge. The fact that bovid representations (clay models, pottery designs, figurations in rock art and in the form of ritual deposits) took a very important place in the early farming societies of Europe, entirely new in comparison to the Mesolithic bestiary (Tresset 2005), points out that not only the animals and the skills attached to them were diffused, but also their symbolic values (Tresset & Vigne 2007), which were already obvious in the Near Eastern PPN Neolithic (Peter & Schmidt 2004). As such, symbolism was certainly a major element in the colonisation process. In the course of the diffusion of bovids, it appears that interaction with native Mesolithic societies also took place.

Interestingly, data recently published on the appearance and early development of pig husbandry in Europe (Larson *et al.* 2007) suggest a scenario very different from the one we depict for bovids. Larson *et al.* have shown on the basis of a very large study of both ancient and modern DNA that a first Neolithic introduction of pigs from the Near East, very similar to the one observed for bovids, was shortly followed by a local domestication of the European wild boar. This phenomenon was of such a magnitude that, at the end of the Neolithic, very few of the lineages initially introduced from the Near East survived and they had all disappeared by the Bronze Age. It can be hypothesised that both the successful adaptation of Near Eastern domesticates to completely new environments at the start of the European Neolithic and the subsequent interactions with Mesolithic hunters during the diffusion process of these animals had created the conceptual conditions of an entirely European creation: pig husbandry. But this is another story...

Acknowledgements

Results synthesised in this paper have benefitted from various founded projects: “*Early domestic bovids in Europe and the Near-East*” (dir. J.-D. Vigne and M. Bruford; 2000–2003) in the frame of the OMLL scheme (EUROCORES, ESF); “*Premières diffusions des bovidés domestiques au Proche-Orient et en Europe*” (dir. J.-D. Vigne, 2004–2005) in the frame of the ACI OHLL (French Ministry for Research); and “*Chronobos*” (dir. S. Hughes, 2005–2008) in the frame of the ANR “*Genanimal*” (project number: ANR -GANI-004). R. Bollongino held a post-doctoral position (2006–2007) funded by the CNRS in the field of palaeogenetics of early domestic bovines. C.J. Edwards held two post-doctoral positions (2002–2004, and 2004–

2007) at Trinity College, Dublin (TCD), Ireland, both supported by the Irish Research Council for Science Engineering and Technology (project numbers SC/1999/409 and SC/2002/510). Work at TCD was also part-funded by The Wellcome Trust (grant no. 047485/Z/96/Z).

References

- Achilli, A., Olivieri, A., Pellecchia, M., Ubaldi, C., Colli, L., Al-Zahery, N., Accetturo, M., Pala, M., Kashani, B.H., Perego, U.-A., Battaglia, V., Fornarino, S., Kalamati, J., Houshmand, M., Negrini, R., Semino, O., Richards, M., Macaulay, V., Ferretti, L., Bandelt, H.-J., Ajmone-Marsan, P. & Torroni, A. 2008. Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Current Biology* 18, R157–8.
- Arbuckle, B.S. 2005. Experimental animal domestication and its application to the study of animal exploitation in prehistory, in *New Methods for the First Steps of the Domestication*, eds. J.-D. Vigne, D. Helmer & J. Peters. Oxford: Oxbow Books, 18–33.
- Balasse, M. & Ambrose, S.H. 2005. Distinguishing sheep and goats using dental morphology and stable carbon isotopes in C4 grassland environments. *Journal of Archaeological Science* 32, 691–702.
- Balasse, M. & Tresset, A. 2007. Environmental constraints on reproductive activity of domestic sheep and cattle: what latitude for the herder? *Anthropozoologica* 42, 71–88.
- Balasse, M., Tresset, A. & Ambrose, S.H. 2006. First evidence for seaweed winter foddering in the Neolithic of Scotland. *Journal of Zoology* 270, 170–6.
- Bollongino, R. & Vigne, J.-D. 2007. Temperature monitoring in archaeological animal bone samples in the Near East semi arid area, during and after the excavation. *Journal of Archaeological Science* 35, 873–81.
- Bollongino, R., Edwards, C.J., Burger, J., Alt, K.W. & Bradley, D.G. 2006. Early history of European domestic cattle as revealed by ancient DNA, *Biology Letters* 2, 155–9.
- Bollongino, R., Tresset, A., Vigne, J.-D. 2008a. Environment and excavation: Pre-lab impacts on ancient DNA analyses. *Comptes Rendus Palevol* 7, 91–8.
- Bollongino, R., Elsner, J., Vigne, J.-D. & Burger, J. 2008b. Y-SNPs do not indicate hybridisation between European aurochs and domestic cattle. *Public Library of Science One* 3, e3418.
- Degerbøl, M. & Fredskild, B. 1970. *The Urus (Bos primigenius Bojanus) and Neolithic domesticated cattle (Bos taurus domesticus Linné) in Denmark*. Copenhagen: The Royal Danish Academy of Sciences and Letters.
- Digard, J.-P. 1990. *L'Homme et Les Animaux Domestiques. Anthropologie d'une Passion*. Paris: Fayard.
- Edwards, C.J., Machugh, D.E., Dobney, K.M., Martin, L., Russell, N., Horwitz, L.K., McIntosh, S.K., Macdonald, K.C., Helmer, D., Tresset, A., Vigne, J.-D. & Bradley, D.G. 2004. Ancient DNA analysis of 101 cattle remains: limits and prospects. *Journal of Archaeological Science* 31, 695–710.
- Edwards, C.J., Bollongino, R., Scheu, A., Chamberlain, A., Tresset, A., Vigne, J.-D., Baird, J.F., Larson, G., Ho, S.Y.W., Heupink, T.H., Shapiro, B., Freeman, A.R., Thomas, M.G., Arbogast, R.-M., Arndt, B., Bartosiewicz, L., Benecke, N., Budja, M., Chaix, L., Choyke, A.M., Coquegniot, E., Döhle, H.-J., Göldner, H., Hartz, S., Helmer, D., Herzig, B., Hongo, H.,

- Mashkour, M., Özdoğan, M., Pucher, E., Roth, G., Schade-Lindig, S., Schmölcke, U., Schulting, R., Stephan, E., Uerpmann, H.-P., Vörös, I., Voytek, B., Bradley, D.G. & Burger, J. 2007. Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proceedings of the Royal Society Series B* 274, 1377–85.
- Fernández, H., Hughes, S., Vigne, J.-D., Helmer, D., Hodgins, G., Miquel, C., Hanni, C., Luikart, G. & Taberlet, P. 2006. Divergent mtDNA lineages of goat in an early Neolithic site, far from the initial domestication areas. *Proceedings of the National Academy of Sciences of the U.S.A* 103, 15375–9.
- Geigl, E.-M. 2008. Palaeogenetics of cattle domestication in Southwest Asia: Methodological challenges for fossil bones preserved in hot regions. *Palevol* 7, 99–112.
- Götherström, A., Anderung, C., Hellborg, L., Elburg, R., Smith, C., Bradley, D.G. & Ellegren, H. 2005. Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Proceedings of the Royal Society Series B* 272, 2345–50.
- Guilaine, J., Manen, C. & Vigne, J.-D. (eds.) 2007. *Pont de Roque-Haute (Portiragnes, Hérault). Nouveaux Regards sur la Néolithisation de la France Méditerranéenne*. Toulouse: Centre d'Anthropologie (Archives d'Ecologie Préhistorique), 332.
- Halstead, P., Collins, P. & Isaakidou, V. 2002. Sorting the sheep from the goats: morphological distinctions between the mandibles and the mandibular teeth of adult *Ovis* and *Capra*. *Journal of Archaeological Science* 29, 545–53.
- Harris, D.R. (ed.) 1996. *The Origins and Spread of Agriculture and Pastoralism in Eurasia*. Washington DC: Smithsonian Institution.
- Helmer, D. 2000. Discrimination des genres *Ovis* et *Capra* à l'aide des prémolaires inférieures 3 et 4 et interprétation des âges d'abattages: l'exemple de Dikili Tash (Grèce). *Anthropozoologica* 31, 29–38.
- Helmer, D. & Gourichon, L. in press. Premières données sur les modalités de subsistance dans les niveaux récents (PPNB moyen à Néolithique à Poterie) de Tell Aswad en Damascène (Syrie), Fouilles 2001–2005, in *Archaeozoology of the Near East VIII, Proceedings of the 4th International Symposium on Archaeozoology of Southwestern Asia and Adjacent Areas*, eds. E. Vila & L. Gourichon. Lyon: Maison de l'Orient Méditerranéen.
- Helmer, D., Gourichon, L. & Stordeur, D. 2004. A l'aube de la domestication animale. Imaginaire et symbolisme animal dans les premières sociétés Néolithiques du nord du Proche-orient. *Anthropozoologica* 39, 143–64.
- Helmer, D., Gourichon, L., Monchot, H., Peters, J. & Saña Seguí, M. 2005. Identifying early domestic cattle from Pre-Pottery Neolithic sites on the Middle Euphrates using sexual dimorphism, in *First Steps of Animal Domestication, New Archaeozoological Approaches*, eds. J.-D. Vigne, J. Peters & D. Helmer. Oxford: Oxbow Books, 86–95.
- Ho, S.Y., Larson, G., Edwards, C.J., Heupink, T.H., Lakin, K.E., Holland, P.W. & Shapiro, B. 2008. Correlating Bayesian date estimates with climatic events and domestication using a bovine case study. *Biology Letters* 4, 370–4.
- Larson, G., Albarella, U., Dobney, K., Rowley-Conwy, P., Schibler, J., Tresset, A., Vigne, J.-D., Edwards, C.J., Schlumbaum, A., Dinu, A., Balasescu, A., Dolman, G., Tagliacozzo, A., Manaseryan, N., Miracle, P., Van Wijngaarden-Bakker, L., Masseti, M., Bradley, D.G. & Cooper, A. 2007. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences of the U.S.A* 104, 15276–81.
- Legge, A.J. 1972. Prehistoric exploitation of the gazelle in Palestine, in *Papers in Economic Prehistory*, ed. E.S. Higgs. Cambridge: Cambridge University Press, 119–24.

- Meadow, R.H. 1984. Animal domestication in the Middle East: a view from the eastern margin, in *Animals and Archaeology 3. Early Herders and their Flocks*, eds. J. Clutton-Brock, G. Grigson & D.R. Harris. Oxford: British Archaeological Reports S202, 309–37.
- Naderi, S., Rezaei, H.R., Taberlet, P., Zundel, S., Rafat, S.A., Naghash, H.R., El-Barody, M.A., Ertugrul, O., Pompanon, F. & Econogene Consortium 2007. Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *Public Library of Science ONE* 2, e1012.
- Naderi, S., Rezaei, H.R., Pompanon, F., Blum, M.G., Negrini, R., Naghash, H.R., Balkiz, O., Mashkour, M., Gaggiotti, O.E., Ajmone-Marsan, P., Kence, A., Vigne, J.-D. & Taberlet, P. 2008. The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences of the U.S.A* 105, 17659–64.
- Nielsen-Marsh, C.M. & Hedges, R. 2000a. Patterns of diagenesis in bone I: The effects of site environments. *Journal of Archaeological Science* 27, 1139–50.
- Nielsen-Marsh, C.M. & Hedges, R. 2000b. Patterns of diagenesis in bone II: Effects of acetic acid treatment and the removal of diagenetic CO²⁻³. *Journal of Archaeological Science* 27, 1151–9.
- Pascal, M., Lorvelec, O. & Vigne, J.-D. 2005. Holocene turnover of the French vertebrate fauna. *Biological Invasions* 7, 99–106.
- Pascal, M., Lorvelec, O. & Vigne, J.-D. 2006. Holocene turnover of the French vertebrate fauna: Corrigendum. *Biological Invasions* 8, 395.
- Peters, J. & Schmidt, K. 2004. Animal and the symbolic world of Pre Pottery Neolithic Göbekli Tepe, south eastern Turkey: preliminary assessment. *Anthropozoologica* 39, 179–218.
- Peters, J., Helmer, D., Von Den Driesch, A. & Saña Segui, M. 1999. Early animal husbandry in the Northern Levant. *Paléorient* 25, 27–47.
- Peters, J., Von Den Driesch, A. & Helmer, D. 2005. The upper Euphrates-Tigris basin: cradle of agro-pastoralism? in *First Steps of Animal Domestication, New Archaeozoological Approaches*, eds. J.-D. Vigne, J. Peters & D. Helmer. Oxford: Oxbow Books, 96–124.
- Poplin, F. 1979. Origine du mouflon de Corse dans une nouvelle perspective paléontologique: par marronnage. *Annales Génétique et Sélection Animale* 11, 133–43.
- Poplin, F., Poulain, T., Meniel, P., Vigne, J.-D., Geddes, D. & Helmer, D. 1986. Les débuts de l'élevage en France, in *Le Néolithique de la France (Hommage à G. Bailloud)*, eds. J.-P. Demoule & J. Guilaine. Paris: Picard Édition, 37–51.
- Redding, R. 2005. Breaking the mould: a consideration of variation in the evolution in animal domestication, in *First Steps of Animal Domestication: New Archaeozoological Approaches*, eds. J.-D. Vigne, J. Peters & D. Helmer. Oxford: Oxbow Books, 41–8.
- Scheu, A., Hartz, S., Schmölckle, U., Tresset, A., Burger, J., Bollongino, R. 2007. Ancient DNA provides no evidence for independent domestication of cattle in Mesolithic Rosenhof, Northern Germany. *Journal of Archaeological Science* 35, 1257–64.
- Smith, C.I., Chamberlain, A.T., Riley, M.S., Stringer, C. & Collins, M.J. 2003. The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of Human Evolution* 45, 203–17.
- Svensson, E. & Götherström, A. 2008. Temporal fluctuations of Y-chromosomal variation in *Bos taurus*. *Biology Letters* Aug 19th [Epub ahead of print].
- Tanno, K. & Willcox, G. 2005. How fast was wild wheat domesticated? *Science* 311, 1886.
- Tresset, A. 2000. Early husbandry in Atlantic areas. Animal introductions, diffusion of techniques and native acculturation at the north-western margin of Europe, in *The Prehistory*

- and *Early History of Atlantic Europe*, ed. J. Henderson. Oxford: British Archaeological Reports International Series, 17–32.
- Tresset, A. 2003. French connections II: Of cows and men, in *Neolithic Settlement in Ireland and Western Britain. Actes du Colloque de la Prehistoric Society, Belfast, Avril 2001*, eds. I. Armit, E. Murphy, E. Nelis. & D. Simpson. Oxford: Oxbow Books, 18–30.
- Tresset, A. 2005. La place changeante des bovin dans les bestiaires du Mésolithique final et du Néolithique du Nord-Ouest européen in *Unité et Diversité du Processus de Néolithisation de la Façade Atlantique de l'Europe (7^e-4^e Millénaires avant notre Ère). Mémoire de la Société Préhistorique Française*, eds. G. Marchand & A. Tresset. **36**: 271–86.
- Tresset, A. & Vigne, J.-D. 2006. Le dépôt d'animaux de la structure e4. Une illustration de la symbolique des bovins à la charnière du Mésolithique et du Néolithique bretons? in *Monuments Mégalithiques à Locmariaquer (Morbihan). Le Long Tumulus d'Er Grah dans son Environnement*, ed. C.-T. Le Roux. Paris: CNRS éd. (38^e suppl. *Gallia Préhistoire*), 123–45.
- Tresset, A. & Vigne, J.-D. 2007. Substitution of species, techniques and symbols at the Mesolithic/Neolithic transition in Western Europe, in *Going Over: the Mesolithic/Neolithic Transition in NW Europe*, eds. A. Whittle & V. Cummings. London: Proceedings of the British Academy 144, 189–210.
- Troy, C.S., MacHugh, D.E., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., Chamberlain, A.T., Sykes, B.C. & Bradley, D.G. 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410, 1088–91.
- Uerpmann, H.-P. 1979. *Probleme der Neolithisierung des Mittelmeerraums*. Wiesbaden: Dr Ludwig Reichert Verlag, 190.
- Van Vuure, C. 2005. *Retracing the Aurochs: History, Morphology and Ecology of an Extinct Wild Ox*. Sofia, Pensoft.
- Vigne, J.-D. 2008. Zooarchaeological aspects of the Neolithic diet transition in the Near East and Europe, and their putative relationships with the Neolithic demographic transition, in *The Neolithic Demographic Transition and its Consequences*, eds. J.-P. Bocquet Appel & O. Bar-Yosef. New York: Springer Verlag, 179–205.
- Vigne, J.-D. in press. Introduction et réintroduction de mammifères à Chypre aux 9^e et 8^e millénaires av. J.-C. (Néolithique précéramique): indices indirects de l'usage de la voile au Néolithique? in *Hommages à Jean Guilaine*, eds. B. Midens-Reyne & C. Manen. Toulouse: Centre d'Anthropologie.
- Vigne, J.-D. & Helmer, D. 1999. Nouvelles analyses sur les débuts de l'élevage dans le Centre et l'Ouest méditerranéens, in *Le Néolithique du Nord-Ouest Méditerranéen. Actes XXIVe Congr. Préhist. Fr., Carcassonne (26–30 September 1994)*, 129–46.
- Vigne, J.-D., Carrère, I., Saliege, J.-F., Person, A., Bocherens, H., Guilaine, J. & Briois, F. 2000. Predomestic cattle, sheep, goat and pig during the late 9th and the 8th millennium cal. BC on Cyprus: preliminary results of Shillourokambos (Perkklisha, Limassol), in *Archaeozoology of the Near East IV, Proceedings of the 4th International Symposium on Archaeozoology of Southwestern Asia and Adjacent Areas*, eds. M. Mashkour, A.M. Choyke, H. Buitenhuis & F. Poplin. Groningen: Archaeological Research and Consultancy (Publicaties 32), 52–75.
- Vigne, J.-D., Carrère, I. & Guilaine, J. 2003. Unstable status of early domestic ungulates in the Near East: the example of Shillourokambos (Cyprus, IX-VIIIth millennia cal. BC), in *Le Néolithique de Chypre*, eds. J. Guilaine & A. Le Brun. Athens: Ecole Fr. Athènes (*Bull. Corr. Hellenique*, supplement 43), 239–51.
- Vigne, J.-D., Helmer, D. & Peters, J. 2005. New archaeozoological approaches to trace the first steps of animal domestication: general presentation, reflections and proposals, in *First Steps*

- of Animal Domestication: New Archaeozoological Approaches*, eds. J.-D. Vigne, J. Peters & D. Helmer. Oxford: Oxbow Books, 1–16.
- Woodman, P., McCarthy, M. & Monaghan, N. 1997. The Irish Quaternary Fauna Project. *Quaternary Science Reviews* 16, 129–59.
- Zeder, M. 2005. A View from the Zagros: new perspectives on livestock domestication in the Fertile Crescent, in *First Steps of Animal Domestication, New Archaeozoological Approaches*, eds. J.-D. Vigne, J. Peters & D. Helmer. Oxford: Oxbow Books, 125–46.

Language and genes

Linguistic, cultural and genetic perspectives on human diversity in west-central Africa

Lolke Van der Veen¹, Lluís Quintana-Murci² and David Comas³

This chapter presents the major results of a recent multidisciplinary study of the human diversity of west-central Africa. This diversity is examined from linguistic, cultural and population genetic (mtDNA and Y chromosome) perspectives. The study offers new insights into (i) the peopling of the Cameroon-Gabon area (the so-called “Bantu expansion”), (ii) the linguistic, cultural and genetic exchanges between the various groups of Bantu-speaking farmers over the last three millennia, and (iii) the recent and ancient relationships between these populations and Pygmy hunter-gathering groups. Evidence from mtDNA diversity suggests an initial divergence of the ancestors of contemporary Pygmies from an ancestral central African population starting not earlier than ~70,000 years ago. Evidence from both mtDNA and Y-chromosome variation suggests long-standing and asymmetrical gene flow between the two types of populations.

Introduction

Much of the peopling of central Africa and the population movements within this area over the last four or five millennia remain unknown or uncertain to date. Except for the last few centuries AD, written documentation is inexistent. The major contributions to the scholarly study of this period have come from linguistics and archaeology so far. During the 20th century, linguists have developed more or less consensual theories about the so-called Bantu expansion⁴, the major lines of which many specialists now agree on. The archaeological record has also yielded information about the gradual dispersal and settling of villagers practising

-
1. « Dynamique du Langage » Laboratory, CNRS & Université Lumière-Lyon 2, France
 2. UP Human Evolutionary Genetics, Institut Pasteur, Paris, France
 3. Unitat de Biologia Evolutiva, Universitat Pompeu Fabra, Barcelona, Spain
 4. The some 500 Bantu languages form a subgroup of one of the subbranches of the Niger-Congo language family. Geographically, they cover most of sub-Saharan Africa.

a rudimentary form of subsistence agriculture, their probable encounters and exchanges with smaller groups of hunter-gatherers already in situ, and the relationships between the spread of these villagers and technological innovations such as agriculture and iron metallurgy.

The present chapter intends to demonstrate how careful and well-concerted multidisciplinary research may provide new and better insights into the complex nature of the history of human populations in west-central Africa, a region which provides an excellent test case for the languages and genes debate. The spread of the agriculturalists occurred during a period of time with limited time depth, not exceeding four millennia, and which is characterized both by a growing population density that lead to more contact and exchanges between population groups with different lifestyles and by an increase of population movements. Furthermore, linguists from Lyon, Tervuren (Belgium) and Libreville (Gabon) have been carrying out extensive research on the languages varieties of Gabon and its neighbouring regions (Cameroon, parts of Congo-Brazzaville and DRC) over the last two decades, which has led to a better understanding of both the linguistic and cultural diversity of the region and to the elaboration of hypotheses about the historical processes that may have shaped the present-day distributions. West-central Africa attests the highest linguistic diversity for Bantu-speaking Africa. Farming villagers and groups of hunter-gatherers (often abusively referred to as Pygmies⁵) coexist. The latter usually live at the fringe of the forest, and their traditional nomadic lifestyle is rapidly fainting. As a rule, the village communities living north to the Ogooué River generally have a patrilineal descent system, whereas most of the other ethnic groups, located to the south of this river, which runs across the south-eastern part of Gabon and its centre, are matrilineal. Modern lifestyle tends to transform these systems into a more Western-like system. Patrilocality (i.e. the socio-cultural practice whereupon a married couple resides with or near the family of the husband) is the rule. Mating patterns between members of the village communities and the hunter-gathering communities are asymmetrical: male villagers may (and do) marry Pygmy women, but Pygmy men will not marry female villagers in general.

Hypotheses based on the linguistic and cultural distributions within the area may be used to explain the progressive peopling of the region by the Bantu-speaking farmers, especially when oral literature concerning the migration paths and the peoples' origin and the results from thorough comparative linguistic research tend to corroborate each other. According to these hypotheses the Gabon area

5. Although the term will be used hereafter for the sake of convenience, the authors regard the term "Pygmy" as inappropriate because of its pejorative connotations. It refers to a variety of (mostly) hunter-gathering populations, also known as the "Forest peoples".

must have been the scene of successive waves of migrations splitting off from a major southbound western Bantu expansion, penetrating the territory from the north(-east), the south and the south-east⁶. These waves gave rise to quite extensive linguistic and cultural convergence, especially in the central part of Gabon.

In order to improve our understanding of the diversity and the migration history of the region, we engaged into a multidisciplinary study called “Language, culture and genes in Bantu”, which brought together specialists of different disciplines including linguists, ethnologists, geneticists, archaeologists and historians. Some 21 population groups and in most cases at least 50 individuals per population were carefully sampled, following very rigorous criteria to ascertain ethnolinguistic accuracy and representativity. About 1000 blood samples were collected for analysis from these 21 populations in 13 different locations, directly in the field⁷.

Although languages and genes do not necessarily correlate, it is obviously advantageous to marry a person that speaks the same or a related language. Other favourable conditions may be low population density and geographic isolation. As several of these conditions were met we wondered whether languages and genes (populations?) coevolved in this equatorial region among the Bantu-speaking populations. Only a genuinely multidisciplinary approach would be capable of tackling such a highly complex question.

In the first section of this chapter two issues related to the linguistic and cultural diversity in west-central Africa will be addressed: 1. the linguistic situation in relation to the dichotomy farming villagers vs. hunter-gatherers, and 2. the so-called Fang enigma. The latter refers to the Bantu-speaking Fang population (Gabon, Cameroon), many members of which rather surprisingly claim a Sudanic (“Egyptian”) origin on the basis of aspects of their oral tradition.

The next section will give an overview of the project’s results for the analyses of both maternally-inherited mtDNA and paternally-inherited Y-chromosome diversity of the sampled populations (both farming villagers and hunter-gatherers). As indicated before, these analyses will also take into account data that has recently become available, particularly from Cameroon, in order to widen the scope of the study.

The chapter will end with a summary and discussion of the major conclusions as well as some perspectives for future research.

6. According to Rexová et al. (2006) the populations that entered the territory from the south and/or the south-east did not split off from a western expansion movement, but originated from an area west of the Great Lakes.

7. In addition to these core samples, other samples, already available or collected in the framework of other recent research programmes, have been taken into account for both mtDNA and Y-chromosome analyses. This explains why the number of samples and/or populations mentioned in the different parts of this chapter varies.

Aspects of linguistic and cultural diversity in west-central Africa

Farming villagers vs. Pygmy hunter-gatherers

Synchronic linguistic distributions

According to Bastin's *et al.* (1999) lexicostatistically-based classification (fig. 1., Bantu can be subdivided into four major groups of unequal size: the very small Mbam/Bubi group, north-western Bantu, west-central Bantu (central-western hereafter), and eastern and southern Bantu. Holden & Gray (2006), who experimentally applied methods used in phylogenetic analysis to linguistic data, obtained similar patterns within their network structure. Due to the difficulty to disentangle the complex linguistic networks that result from convergence, no satisfactory complete sub-classification has been achieved for Bantu so far. Simple traditional dendrogrammes (trees) are clearly insufficient. Therefore, a broad picture has to be assumed where *Forest Bantu* (roughly Guthrie's A, B and C zones⁸) opposes to *Savannah Bantu* (the remainder). In depth regional studies may contribute to the elaboration of a more comprehensive internal classification.

The Gabon area also attests a highly complex situation (Van der Veen 2006): several languages are spoken in more than one place or region. Figures 2 and 5 only indicate the centres of gravity for each of the approximately 50 language varieties! Within this mosaic of isoglosses a major linguistic boundary appears between north-western and central-western Bantu (cf. Figure 2 and 5), although blurred by convergence phenomena in several places. The north-western vs. central-western dichotomy is based on several criteria such as the number of phonemic vowels, the presence or absence of distinctive vowel length, the presence or absence of obstruent spirantization, shared lexical innovations, shared morphological innovations; and possibly also by specific tone features.

In Gabon, north-western Bantu comprises A zone languages (Benga A34, Fang A75, Shiwa A83 and Bekwil A85b) as well as the MYENE and TSOGO (B10 and B30) groups, with most of the KELE (B20) group. Within north-western Bantu, MYENE and TSOGO together form a distinct subgroup (cf. Mouguiama-Daouda & Van der Veen 2005), sharing several, mainly lexical traits, which may be explained either by inheritance or by convergence due to prolonged contact. The precise status of the very complex and geographically shattered KELE group requires further elucidation (cf. Bastin *et al.* 1999, where B20 emerges as a floating group). Alewijnse *et al.* (2007), who provide a new (tentative) computationally-based classification for most of the languages of Gabon, confirm its unity but do not yield new evidence concerning its affiliation. In this classification, the B10-B30 cluster

8. Malcolm Guthrie's (1967–71) classification and references are used throughout this chapter.

does not appear as solid as thought previously, which favours the similarity-by-convergence hypothesis. New evidence from ongoing research on KELE suggests KELE may well be an (independent?) ancient branch of north-western Bantu.

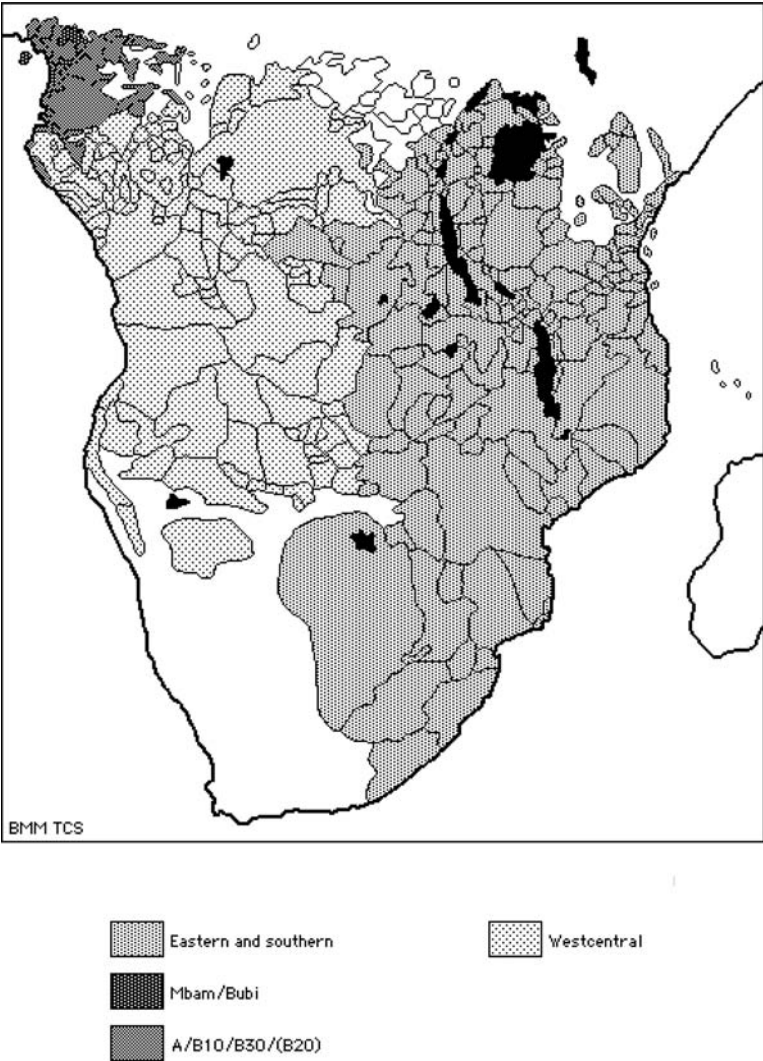


Figure 1. Major groups of Bantu according to Bastin *et al.* 1999: Mbam/Bubi, north-western (A/B10/B30/(B20)), west-central, and eastern and southern. Acknowledgement to Thilo C. Schadeberg for his Bantu Map Maker tool

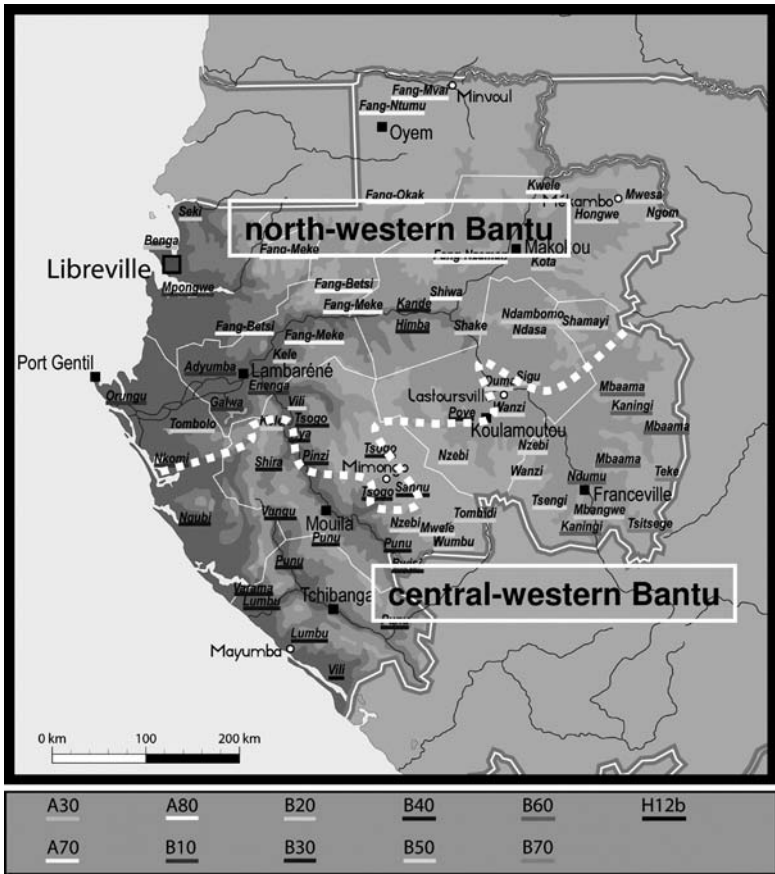


Figure 2. Gabon area: Languages spoken by the villagers. Major linguistic boundary appears within the area, blurred in several places by convergence phenomena. (Map: DDL Africanists. Lyon)

Central-western Bantu in Gabon is represented by the SHIRA–Vili (B40–H12) and the NJABI–MBETE–TEKE (B50–B60–B70) clusters. These clusters are respectively related to H zone and C zone languages. The proximity between B50 and the B60–B70 sub-cluster may also be due to substantial convergence.

In west-central Africa, the farming villagers all speak Bantu language varieties (Figure 2), most of which are closely related.

However, the linguistic situation of the small groups of (semi-)nomadic hunter-gatherers —traditionally called Pygmies or “Forest peoples”— is more heterogeneous (Figure 3). According to Mayer (1987) three different linguistic situations can be distinguished for these populations:

1. Most hunter-gathering groups have adopted a nearby Bantu language and (presumably) abandoned their former language (*language shift*). This most common strategy may well be the sacrifice these groups are willing to make in order to preserve at least part of their cultural identity. Sometimes these non-villagers varieties retain archaic traits no longer attested in the present-day neighbouring language(s).
2. Some groups of (semi-)nomads seem to be using mixed Bantu language varieties made up of elements taken from several surrounding Bantu languages.
3. Finally, some hunter-gatherers speak non-Bantu languages, which may also have been adopted at some earlier stage. E.g. in Gabon, the Bakao speak Baka, an Ubangian language.

Hunter-gatherers are generally looked upon as some inferior, animal-like race by the villagers who often consider them as their personal property. However, paradoxically, the villagers very much admire and appreciate the medical, musical and

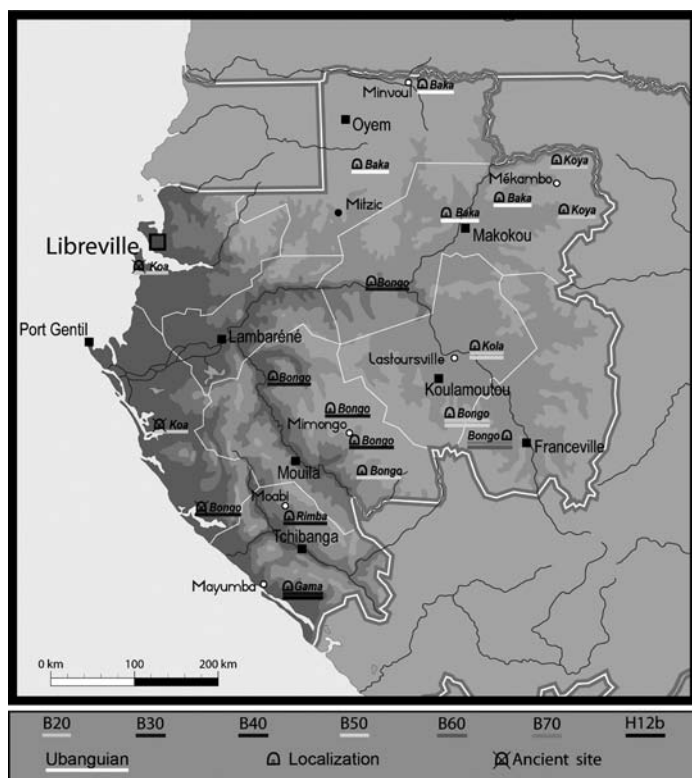


Figure 3. Gabon area: Language varieties spoken by the hunter-gatherer communities. Map: DDL Africanists, Lyon

hunting skills of the Pygmies. There are indications that some of the nomadic groups might still be using their primitive languages, be it in secret and for in group communication only. This is currently being investigated. If such languages still exist, it will be interesting to know whether they also belong to the Niger-Congo language family or to some other phylum, known or unknown.

Processes capable of accounting for the nowadays patterns

Bantu-speaking west-central Africa attests the highest degree of linguistic diversity (at the group level) and is located near the closest linguistic relatives of Bantu. For these reasons, the original Bantu homeland is nowadays considered to be located in the north-western part of this area, i.e. the Bamenda Highlands of Cameroon, near Mount Cameroon. Archaeological evidence supports this point of view (cf. Clist 2005). From here, the dispersal must have been slow, gradual and wave-like, following an initial period of dialectal fragmentation (see Figure 4 for a schematic view). Glottochronology (a controversial approach sometimes used in historical linguistics for estimating the time at which languages diverged) suggests a relative time depth of some 5000 years for the beginning of the diversification process. However, recent evidence from archaeology (Clist 2005) rather pleads for an approximate time depth of 4000 years. Climate conditions were favourable (cf. Maley 2001).

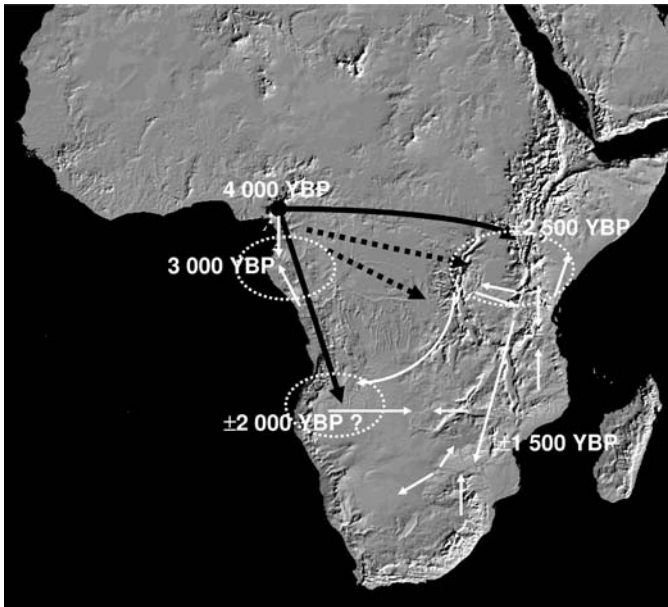


Figure 4. Schematic view of the Bantu “expansion” and relative chronology. Black arrows: major dispersals; white arrows: minor dispersals; dotted-lined ellipses: major linguistic convergence zones. (Map: Van der Veen.)

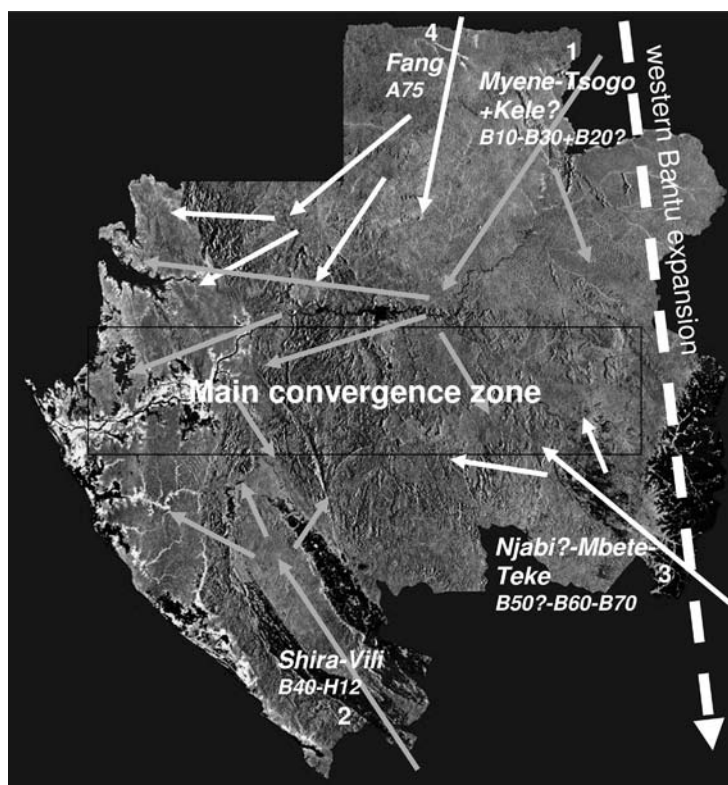


Figure 5. Broad scenario of Bantu languages spreading across the Gabon area. Major dispersal wave (*southbound, white dotted arrow*) and overall migration paths inferred from diachronic language studies and oral literature. Numbers indicate relative chronology. Language reference numbers according to Guthrie (1967–71)

West-central Africa must have been the scenery of the gradual spread of Bantu, in particular the western, southbound, dispersal⁹. The present, often extremely complex, linguistic situations (with numerous dialect continua and prevailing multilingualism) can be accounted for by gradual fragmentation and extensive linguistic convergence (language admixture) due to prolonged contact and multilingualism. It should also be kept in mind that several languages and/or populations probably disappeared in the course of time. The extent of language replacement may have been considerable.

9. Rexová et al. (2006), who propose a new tree based on both lexical and grammatical data, do not really question this western expansion but considers the western languages south of the rainforests to originate from a region west of the Great Lakes, and thus limits the extent of the western expansion. One may regret the absence of several important B zone languages in this stimulating study.

Figure 5 proposes a broad and necessarily oversimplified scenario based on an extensive and careful study of both the distribution of linguistic traits within the area and oral literature, and also on archaeological data and/or the limited available historical data. This study suggests the area was occupied by successive waves of Bantu-speakers coming from the north, the south and the east. Convergence by contact seems widespread, especially in the centre.

Archaeological evidence strongly suggests that when the Neolithic farming villagers (most probably Bantu-speakers) arrived in the area, the hunter-gathering populations had already been *in situ* for a long time. The record provides relatively few signs of (genetic) admixture between the two types of populations, which seem to have lived in peaceful coexistence for more than two millennia at least, engaged in mainly economic exchanges.

The Fang enigma

Fang is the name of a population located in southern Cameroon and most of northern Gabon. Its language, carrying the same name, has been classified as Bantu A75. In Gabon, Fang comprises northern, southern as well as some intermediate varieties (Medjo Mvé 1997). As a dominant language spoken by almost 30% of the inhabitants, it has an increasing impact on the neighbouring languages and is currently replacing several of them, i.a. the Shiwa language (Bantu A83) spoken by the Makina population.

Reliable historical sources and indigenous oral literature locate the origin of population further north. However, the precise geographic origin of this, probably much wider, population movement is still a matter of debate (Adamaoua Plateaux in north Cameroon? Upper-Sangha in the Central African Republic?). The following scenario seems plausible: some 1000 YBP, a group of Bantu-speakers left the Yaunde region —most probably not the starting point of this migration— moving southwards. Part of them, the Proto-Fang, also entered the present-day Gabon area. Their wave-like spread, perhaps triggered or catalyzed by incessant invasions by non-Bantu speakers, intensified during the 17th century AD and came to an end only in the first half of the 20th century. Raids related to slave trade by Muslims groups as well as the perspective of economic exchange with European traders certainly contributed to the final stage of the expansion.

In the early 20th century Rev. Trilles (1912a/b, 1931) claimed an “Egyptian” origin for this population, its language and its culture, on the basis of certain elements of its oral tradition (known as ‘Mvet’) and several arguments referring to linguistic, cultural and phenotypic traits. The adherents of this theory posit that the Plateaux of Bahr-el-Ghazal (Upper-Nile, Kenya/Uganda/Sudan area) are the most probable cradle for the ancestral community of Fang. This theory, which is built on a series of speculative and often random parallels drawn by 19th century explorers

and missionaries and suffers from serious methodological and theoretical weaknesses, has become very popular especially among Black African scholars, but also amongst non-scholars. In Gabon, it is a commonly shared belief that the Fang, as well as their language, are not Bantu. It often takes a strong ideological dimension as it accuses (white) Egyptologists of falsifying ancient History. Cheikh Anta Diop (1979, 1981, 1993) and Théophile Obenga (1985, 1993) are the main representatives of this school of thought. It should be noticed that similar claims have been made by other Bantu-speaking populations (e.g. Basaá A43a and Mboshi C25).

Guthrie (1948, 1967–71), Hombert *et al.* (1989) and, more recently, Medjo Mvé (1997) have unambiguously shown that the Fang dialect cluster, in spite of several rather unusual features at the surface (e.g. 8 vowel system + 3 diphthongs, labiovelars, predominantly closed syllables, palatal nasals in word-final position), presents all the traits of a regular Bantu language. There is absolutely no evidence of a non-Bantu substratum. Important parts of Fang lexicon can be directly and straightforwardly related to the Proto-Bantu reconstructions (Guthrie 1967–71) through regular sound correspondences. The language also possesses a regular Bantu noun class system, a regular Bantu verb system, etc. (Hombert *et al.* 1989; Medjo Mvé 1997; Mba-Nkoghe 2001). The alleged structural parallels (essentially, sound correspondences) between ancient Egyptian and Black African languages lack any scientific basis. They are merely non systematic, randomly chosen, chance similarities. Nowadays languages are being compared with ancient languages, e.g. Mboshi (C25) with ancient Egyptian. Core (i.e. basic) lexicon is poorly represented. Moreover, as for the alleged typological similarities, it is a well known fact that this kind of data is insufficient for proving affiliation.

Many outstanding cultural anthropologists who have been working within the area (e.g. Bekombo-Priso & Laburthe-Tolra 1981) also consider the “Out-of-Egypt” theory to be untenable. None of the alleged cultural traits can be considered to be specifically (ancient) Egyptian or Hamitic¹⁰.

Furthermore, as will be mentioned in the next section, the analysis of the mtDNA and the Y-chromosome variation in both Gabonese and Cameroonian populations does not support a non-Bantu (Semitic) origin for Fang either.

On the basis of the available evidence, the “Out-of-Egypt” theory cannot receive any credit at all. Nevertheless, it leaves us with a fascinating example of how a modern myth came suddenly to life and still strongly determines people’s perception¹¹.

10. These languages or language groups belong to a completely different language family, called Afro-Asiatic (formerly, Hamito-Semitic).

11. For a critical overview of the linguistic, anthropological and historical weaknesses of this theory, see Mouguiama-Daouda (2005). This publication also presents evidence drawn from the study of cultural lexicon related to the realms of flora and fauna.

Analysis of maternal lineages (mtDNA) in west-central African populations

Maternal lineages in Bantu agriculturalists and Pygmy hunter-gatherers

In order to unravel the maternal genetic composition of Bantu-speaking agriculturalists (AGR hereafter) and Pygmy hunter-gatherers (PHG hereafter) in west-central Africa, we analyzed mitochondrial DNA¹² (mtDNA) variation in a total of 1404 individuals from 20 AGR populations (983 individuals) and nine PHG populations (421 individuals) from Gabon, Cameroon, the Central African Republic (CAR) and the Democratic Republic of Congo (DRC) (Quintana-Murci *et al.* 2008). MtDNA variation was characterized by direct sequencing of the HVS-I region (from nucleotide position 16024 to 16383) and by genotyping a set of 33 single nucleotide polymorphisms (SNPs)¹³ from the coding region for accurate resolution into haplogroups¹⁴. Thirty-three haplogroups (i.e. major maternal lineages, in the present section) were identified, 32, 13 and 4 of which were present in AGR, western and eastern PHG populations, respectively. Overall, AGR populations displayed higher levels of haplogroup and sequence diversity than PHG populations (e.g. haplogroup diversity in AGR of 0.92 vs. PHG of 0.56). All genetic diversity indices and neutrality tests showed demographic differences between AGR and PHG populations, with AGR populations showing signs of population growth and PHG populations of small population sizes and strong genetic drift.

In terms of mtDNA haplogroup distribution, AGR populations are more diverse than PHG populations (Quintana-Murci *et al.* 2008). Almost all known sub-Saharan African maternal lineages (Salas *et al.* 2002) were represented among AGR populations, with the exception of L0d and L0k, which are typical of the Khoi and San peoples of southern Africa (Knight *et al.* 2003). The most frequent haplogroups (>5%) observed were L1c1a (20.3%), L2a1 (13.3%), L0a1 (6.5%), L1b (6.4%), L3e2 (5.8%), L3e1 (5.7%) and L3f1b (5.7%) (Figure 6). Interestingly, L1c1a is the only haplogroup studied to show almost exclusive geographic clustering with central Africa (Salas *et al.* 2002; Salas *et al.* 2004; Beleza *et al.* 2005; Batini *et al.* 2007; Quintana-Murci *et al.* 2008). The other lineages display variable frequency and

12. Mitochondrial DNA is characterized by strict maternal inheritance, the absence of recombination and a high mutation rate.

13. A single-nucleotide polymorphism is a DNA sequence variation occurring when a single nucleotide (molecule) – A, T, C, or G – in the genome differs between members of a species.

14. In the study of molecular evolution, a haplogroup is a group of evolutionary closely-related lineages (called haplotypes) sharing a common ancestor with a single nucleotide polymorphism (SNP) mutation. A haplotype is a combination of polymorphic mutations inherited together. Haplogroups can be used to define and classify genetic populations and are often geographically oriented.

diversity patterns between west, east and southeast Africa, indicating only that they originated in the equatorial zone. We investigated the phylogenetic and geographic origin of L1c1a, by (i) complete genome sequencing 27 mtDNA molecules covering the widest possible range of L1c variation, as inferred from HVS-I variation and (ii) studying the frequency distribution and diversity of its ancestral types (Quintana-Murci *et al.* 2008). L1c, with sublineages L1c1-L1c6, had a frequency of 27% in central Africa, decreasing to ~5% in both west and southeast Africa. The contemporary geographic distribution of L1c, including all its internal lineages, is consistent with the early arrival of L1c in central Africa (or a central African origin for this lineage, although this cannot be unambiguously proven), followed by a maturation phase in which L1c diversified (i.e. giving rise to its internal derivatives). The virtual restriction of L1c1a to central Africa provides evidence for an autochthonous origin of this lineage within this region. The coalescence age¹⁵ of L1c1a, estimated at $57,100 \pm 7900$ YBP, is also about ~25,000 years greater than those estimated for the other haplogroups frequent in central African AGR populations (15,800 to 29,400 YBP). Altogether, considering the mtDNA data in this region, the contemporary maternal gene pool of AGR populations from central Africa appears to have resulted from a gradual process in which an initial central African gene pool dominated by L1c lineages was subsequently enriched by the introgression (and/or expansion) of the remaining haplogroups such as the L0a, L2 and L3 haplogroups and their derivatives.

The haplogroup lineages found to be present in the different western PHG populations were essentially a subset of those observed among the AGR populations. The first are essentially dominated by two haplogroups that are derivative of L1c1a: L1c1a1 (53.4%) and L1c1a2 (30.1%) (Figure 6). The remaining 16% of lineages belong either to internal derivatives of L1c or to some haplogroups also observed among the groups of Bantu-speakers. Thus, based on the genotyping of our large population sample, aided by complete mtDNA sequences, we can conclude that ~83.5% of contemporary western PHG are descended from a single maternal ancestor of the autochthonous central African lineage L1c1a. The haplogroup profile of the only group of eastern Pygmies here studied – the Mbuti from the Democratic Republic of Congo – included haplogroups L0a2, L2*, L2a2 and L5 (Figure 6). This maternal profile is qualitatively similar to that of eastern African populations (Salas *et al.* 2002) and very different from that of western PHG, indicating a lack of common maternal ancestry between eastern and western PHG.

15. Date at which a particular lineage originated or started to expand demographically.

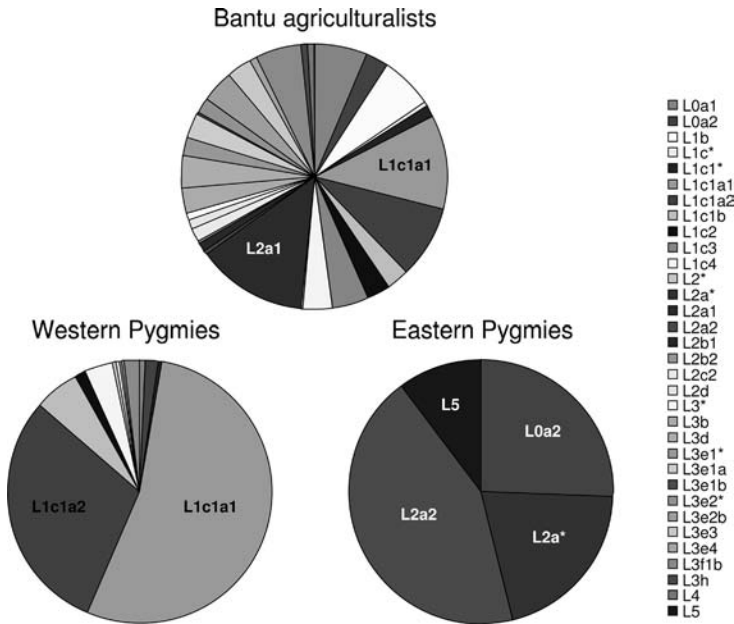


Figure 6.

MtDNA population structure

The genetic structure of these populations was unraveled by means of analyses of the molecular variance (AMOVA method) based on haplogroup distribution (Quintana-Murci *et al.* 2008). When agriculturalists and Pygmy hunter-gatherers were considered as a single group, a certain genetic heterogeneity between the populations analyzed was observed (8%). However, when Bantu-speaking agriculturalists and Pygmies were analyzed independently, the agriculturalists were found to form a very homogeneous group of populations (i.e. only 1.5% of the total genetic variation was due to among-population differences). In addition, estimates of population differentiation (F_{ST}) based on haplogroup frequencies showed that ~60% of inter-population comparisons were not significant. The lack of maternal population differentiation depicted in the group of 20 AGR populations is illustrated in the scatterplot of the first two principal components (PC), on which the AGR populations are tightly clustered (Figure 7). Thus, AGR populations from central Africa displayed a diverse homogeneous pattern with little apparent internal structure. In stark contrast, PHG populations exhibited a high degree of genetic differentiation among them – the AMOVA method revealed that 49% of the total genetic variation was due to among-population differences. The large amount

of among-population diversity observed in the group of Pygmy hunter-gatherers was essentially due to the differences in haplogroup profiles observed between the group of 8 populations of western Pygmies with respect to the Mbuti population of eastern Pygmies. This net separation between western and eastern Pygmies is supported by (i) the analyses of population differentiation based on F_{ST} – all F_{ST} values for pairwise comparisons of eastern Mbuti with western Pygmy populations were very high (0.27–0.47) and highly significant, and (ii) the outlier position of the eastern Mbuti Pygmies on the PC plot (Figure 7). An interesting position in the PC plot involved the Babongo Pygmies – a group of western Pygmies from Gabon. Their position clustering within the AGR group suggests greater gene flow between the Babongo Pygmies and the AGR populations. Interestingly ethnological data indicate that intercultural marriages (in both directions) between Babongo Pygmies and neighbouring farmers are more common in the case of the Babongo (in contrast with other groups of Pygmies), accounting for the greater diversity of this population and its position on the PC plot (Figure 7).

In conclusion, our data show that Bantu-speaking populations from west-central Africa form a very homogeneously related group, at least from the maternal side. This is most probably due to the recent Bantu expansion that might have homogenized the genetic landscape in this region. Despite differences in diversity and demography between AGR and western PHG populations, both groups have the same most common haplogroup (i.e. L1c1a, Figure 6). Haplogroup L1c1a, in the form of the sister lineages L1c1a1 and L1c1a2, has a frequency of 20.3% in AGR populations and 83.5% in western PHG populations. The sharing of the most

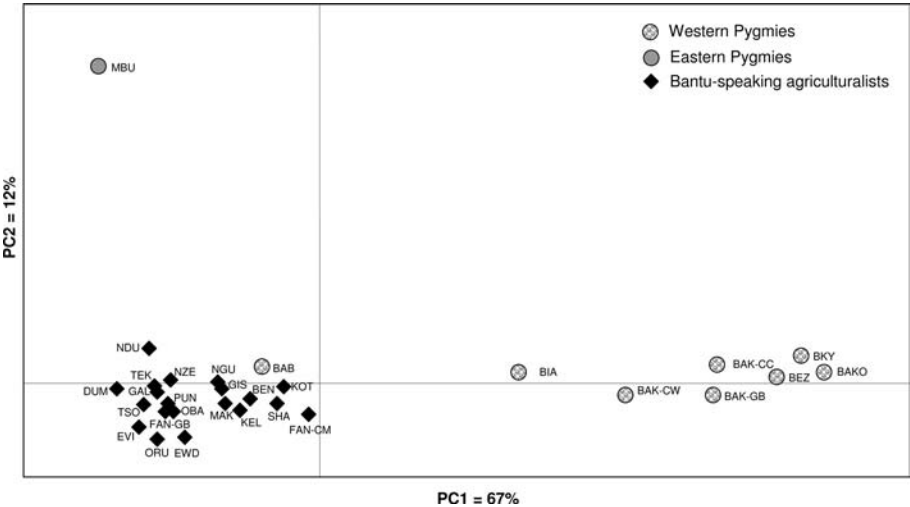


Figure 7.

frequent haplogroup – as well as HVS-I-based haplotypes – between agriculturalists and western Pygmies clearly indicates common maternal ancestry between the two groups and/or high levels of gene flow between them. Finally, the completely different haplogroup profile (both quantitatively and qualitatively) between the group of western Pygmy populations and that of eastern Pygmies attest for a lack of common maternal ancestry between these two groups of PHG populations. However, the extent to which this observation reflects a genuine general lack of common ancestry remains to be determined. Data for 80 autosomal loci¹⁶ from different African populations showed the western Pygmy Biaka and the eastern Pygmy Mbuti to be closely related, with both these groups more similar to west African than to east African populations (Tishkoff & Kidd 2004). More recently, we have conducted a sequence-based study of 20 genome-wide autosomal regions to determine the mode and timing of divergence between western and eastern Pygmies. This study has revealed an early divergence of the ancestors of PHG and AGR populations ~60,000 years ago, a date that is in perfect agreement with that inferred from the maternally-inherited mtDNA. In addition, our autosomal data indicate that the split of the Pygmies' ancestors into the western and eastern PHG groups occurred ~20,000 years ago (Patin *et al.* 2009).

Analysis of paternal lineages (Y chromosome) in west-central African populations

Paternal lineages in Bantu agriculturalists and Pygmy hunter-gatherers

In order to unravel the paternal genetic composition of Bantu-speaking agriculturalists (AGR) and Pygmy hunter-gatherers (PHG) in west-central Africa, we have analyzed a large amount of genetic markers on the Y chromosome in 883 individuals from this region (Berniell-Lee *et al.* 2009). On the whole, most the samples belonged to previously described African lineages (haplogroups), especially common in sub-Saharan Africa. Most of these lineages have been associated to Bantu-speaking AGR populations, such as E1b1a, B2a, and E2, or to PHG populations, such as B2b. Traces of haplogroups A, E*, E1a, and E1b1b1a, found at low frequencies across Africa (Underhill *et al.* 2001; Cruciani *et al.* 2002; Wood *et al.* 2005), were also found in the present sample set. Interestingly, almost 5% of the individuals analyzed belonged to the Eurasian haplogroup R1b1*.

When the paternal lineage composition of AGR and PHG individuals was compared, substantial differences were found (Figure 8). Haplogroups E1b1a and B2a, described as markers of the Bantu expansion (Scozzari *et al.* 1999; Underhill *et al.* 2001; Luis *et al.* 2004; Wood *et al.* 2005; Belezza *et al.* 2005), were the most

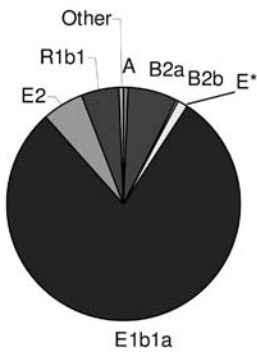
16. Autosomes are chromosomes that are not sex-linked chromosomes.

frequent lineages in Bantu-speaking agriculturalists reaching frequencies of around 80% and 7% respectively, whereas their frequencies in Pygmy hunter-gatherers, especially for E1b1a, were significantly lower (28% and 6%, respectively). In contrast, B2b was present in all the PHG populations of our sample set, reaching a frequency of 47%, but was only found in four AGR individuals (0.5%). The presence of B2b in hunter-gatherers suggests a common origin of paternal lineages in eastern and western Pygmies, which is not found for maternal lineages: eastern Pygmies (Biaka and Mbuti) present L0a2, L2a and L5 maternal lineages and do not present L1c mtDNA lineages, which are found at a frequency of over 90% in western Pygmies (Quintana-Murci *et al.* 2008). Sex-specific demographic factors could explain this discrepancy in maternal and paternal lineages in Pygmies. A remarkable finding is the considerable number of individuals in our sample set who belonged to haplogroup R1b1*, which has been claimed to be a sign of a possible backflow from west Asia into Africa (Cruciani *et al.* 2002).

The paternal lineage composition in AGR and PHG samples in west-central Africa agrees with a homogenizing expansion of Bantu-speakers ancestors that isolated and fragmented Pygmy hunter-gatherers. Bantu-speaking agriculturalists present reduced haplogroup diversity compared to hunter-gathering Pygmies (Figure 8).

Given the composition of the paternal lineages, we can conclude that the genetic exchange (gene flow) between west-central Bantu-speaking agriculturalists and Pygmies has been clearly asymmetrical: paternal Bantu lineages have been introduced into Pygmies, whereas the opposite has been rare. The predominant Bantu haplogroup E1b1a has been found in a frequency of over 25% in Pygmies. On the contrary, the predominant Pygmy haplogroup B2b has been found in less than 1% (4 out of 823) of Bantu-speaking individuals. This result is consistent with the hypothesis of an asymmetrical gene flow between Pygmies and Bantu-speakers due to sex-specific demographic factors (Cavalli-Sforza 1986; Destro-Bisol *et al.* 2004): a maternal Pygmy-to-Bantu flow and a paternal Bantu-to-Pygmy gene flow have been described as the rule. This has also been corroborated by the analyses of female lineages (Batini *et al.* 2007; Quintana-Murci *et al.* 2008), where specific Pygmy mtDNA haplogroups, such as L1c1a, have been found in Bantu-speaking populations, with the opposite situation being rare. Besides this asymmetrical and opposite gene flow in uniparental lineages, it has been shown that there is a substantial common old maternal ancestry in Bantu agriculturalists and Pygmies for the mtDNA (L1c lineages dated back to 70,000 years ago; Batini *et al.* 2007; Quintana-Murci *et al.* 2008). This has not been found for the Y-chromosome: no ancient common paternal lineages have been found in west-central African samples. This lack of ancient paternal lineages in west-central Africa suggests that the “Bantu expansion” erased most of the diversity found in the area before the massive demic expansion.

Bantu agriculturalists



Pygmy

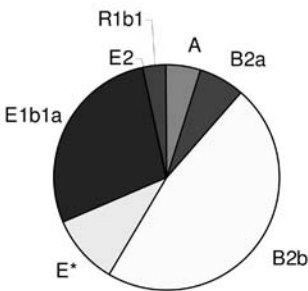


Figure 8.

Y-chromosome population structure

The genetic structure of these populations was unraveled by means of analyses of the molecular variance (AMOVA method) based on lineage composition (Figure 9). When Bantu-speaking agriculturalists and Pygmy hunter-gatherers were considered as a single group, a certain genetic heterogeneity between the samples analyzed was found. However, when Bantu agriculturalists and hunter-gatherers were analyzed independently, the heterogeneity previously found was due to differences among Pygmy populations (19.3%), with Bantu-speaking agriculturalists being

	Between groups	Within groups	Among individuals
All samples		10.8%	89.2%
PHG		19.3%	80.7%
Bantu-speaking AGR		2.3%	97.7%
PHG vs. Bantu-speaking AGR	37.6%	2.3%	60.1%

Figure 9. Analysis of the Molecular Variance (AMOVA). All values are significantly different from zero ($p < 0.05$)

very homogeneous genetically (2.3%). In addition, a large proportion of the total Y-chromosome diversity observed (37.6%) was due to differences between AGR and PHG samples, showing a high genetic difference between these two groups.

In order to establish the genetic relationships between west-central African samples and the rest of the sub-Saharan continent, a Correspondence Analysis (CA) was performed (Figure 10). This analysis aims to reduce all the genetic information obtained in few variables. Thus, the representation of these variables that summarize most of the information in a plot gives an idea of the genetic similarities of the populations analyzed. The representation of the first and second dimensions of the central African groups clusters the populations according to geographic and linguistic proximities with few exceptions. The first axis separates Afro-Asiatic populations from the rest of the populations, characterized by high frequencies of lineages E1b1b, K, T, and F(xG,I,K). However, the scattered distribution of these populations within the Afro-Asiatic cluster shows them to be highly heterogeneous genetically. On the opposite edge of the first axis, the spatial distribution of the (Bantu and non-Bantu) Niger-Congo populations most likely reflects the high frequency of lineages E1b1a and B2a, with no clear differences between Bantu and non-Bantu speakers being appreciated. However, a few Bantu populations (i.e. Fang (Gabon and Cameroon), Punu, etc.) are scattered in the plot due to the presence of haplogroup R1b1*. The second axis locates Khoisan, Pygmy and Nilo-Saharan samples on one edge due to their high frequencies of lineages B2b and A.

As a conclusion, the data presented demonstrate a recent origin for most paternal lineages in west-central Africa as a result of the “Bantu expansion” that erased the Y-chromosome diversity previously found. However, some traces of ancient paternal lineages are found, mainly in hunter-gatherers. These results contrast with the data provided by mtDNA, where ancient lineages are found and substantial maternal gene flow from hunter-gatherers to Bantu-farmers has been suggested. The presence of lineages belonging to R1b1* haplogroup, might represent footprints of demographic expansions in central Africa not directly related to the “Bantu expansion”.

dissimilar. One may infer from this lack of congruence a considerable extent of admixture between the farming populations, and thus an incessant reshaping of the population landscape, encouraged by the geographic and linguistic proximity between the various groups as well as by other cultural, environmental and political factors. The genetic distribution lumps together communities, such as Fang and Teke, who are clearly differentiated at the levels of languages, oral and/or cultural traditions, and also by their position in space. This proximity still lacks a satisfactory explanation, unless one assumes that exchanges were far more developed in former days.

Just one more or less clear grouping, i.e. the MYENE-TSOGO (B10-B30) cluster, emerges from the analysis of the mtDNA variation. However, such a cluster is absent from the study of Y-chromosome diversity, which yields a strong homogeneity between all Bantu samples, without clear clustering or groupings. This may be due to substantial male genetic exchanges between Bantu populations, thus producing increasing homogeneity in male lineages.

The lack of unambiguous correlations between languages and genetic markers suggests that the extent of language replacement has been seriously underestimated for the area. The extremely complex and probably frequent interactions between the different Bantu-speaking populations will most certainly have caused language shift, language merger and language death. The prevailing multilingualism, linguistic proximity, marriage strategies (endogamy, exogamy; polygyny, monogyny; patrilocality, matrilocality), war and local slavery must have played a considerable role here. The composition of ethnolinguistic communities evolved considerably over time, though the integration and loss of individuals. Ongoing research on the languages of the area confirms the existence of numerous dialect chains, causing obvious delimitation problems. As variation is ubiquitous, research is now focusing on centres of gravity.

One may question the reliability of the existing linguistic classifications, nearly exclusively based on lexical data. In depth study of noun and verb morphology is needed, and there are indications that a thorough comparison of the verbal systems of these languages may shed new light on their internal classification. However, it is far from certain that such (necessary) refinements will end up yielding better correlations.

Farming villagers vs. semi-nomadic hunter-gatherers

Although the LCGB project did not initially intend to include hunter-gathering populations, the success of the fieldwork carried out in the Gabon area eventually opened up this opportunity. As results from similar projects, focusing on hunter-gathering

populations of Cameroon and Gabon, became gradually available, the comparison of populations with different lifestyles, traditions and habitats was undertaken.

Agriculturalists and Pygmy hunter-gatherer groups attest significant genetic differentiation, implying both populations coexisted without giving rise to extensive admixture. The differences may have been amplified by the geographic and cultural isolation of the nomads. Nevertheless, a shared mtDNA lineage (L1c, which almost certainly originated in central Africa) suggests an ancient common maternal ancestry between the modern agriculturalists and the western hunter-gatherers. The mtDNA analysis also reveals that western Pygmies, with their common maternal gene pool L1c1a1 and L1c1a2, are clearly distinct from Mbuti hunter-gatherers, the latter being more related to east African populations.

More generally, mtDNA variation suggests that these two groups, which have had very different lifestyles over the last four millennia at least, arose through a complex evolutionary process characterized by (i) initial divergence of the ancestors of contemporary Pygmies from an ancestral central African population no more than approximately 70,000 YBP, (ii) a period of isolation between the two groups, accounting for their phenotypic differences, (iii) long-standing asymmetric maternal gene flow from Pygmies to the ancestors of the farming populations, beginning no earlier than 40,000 YBP and persisting until a few thousand years ago, and (iv) an enrichment of the maternal gene pool of the ancestors of the farming populations by the arrival and/or subsequent demographic expansion of L0a, L2, and L3 carriers (Quintana-Murci *et al.* 2008).

The study of Y-chromosome diversity offers a rather different picture. All Pygmy samples are very heterogeneous, but do share a common ancestral B2b Y-chromosome haplogroup. The presence of both the modal Bantu haplotype and the haplogroups associated to the Bantu expansion (E1b1a and B2a) strongly suggests a Bantu to Pygmy flow of paternal lineages and thus the introduction of typical male villager lineages into the hunter-gatherers' genetic pool. Gene flow in the opposite direction (hunter-gatherers → villagers) is very unlikely, culturally.

Overall, the comparison of mtDNA and Y-chromosome diversity suggests extreme sexually-asymmetrical mating patterns between Pygmies and Bantu-speaking farmers.

Finally, the homogeneity of the Pygmy populations and the shared ancestor of both agriculturalists and hunter-gatherers considerably reinforce the plausibility of the idea that the Pygmy populations, who all speak different language varieties nowadays, abandoned their primitive language(s) in the course of time. The extensive study of specialized lexicon in the present-day languages may still unveil certain properties and elements of these former languages.

Fang enigma

The genetic evidence does not support a Semitic origin for the Fang. However, as for the Y-chromosome variation, this population attests a high frequency of the non-African haplogroup R1b1*. Nonetheless, several other Bantu populations also show the non-African haplogroup, be it with slightly lower frequencies. The exact origin of this lineage requires further investigation. Yet, these results are no proof of a Semitic origin, as the expansion of this paternal lineage is much older (more than 7000 YBP) than the Bantu expansion. It could perhaps be related to the expansion of Nilo-Saharan, another major African language family. In addition, this unusual haplogroup is also found, to various degrees, in some of the other agriculturalist populations examined to date. These populations more or less differ in their oral traditions, their position in space, some of their cultural traits and the presumed time depth of arrival (cf. Figure 5).

What may have caused the introduction of this lineage into these populations? Where and when did this occur? And for what reason(s)? Admixture? Chance? Differences in lifestyle (food, sanitary conditions, etc.)? How did it spread? A variety of historical scenarios may be imagined. Fang and the other R1b1*-carrying populations, such as Teke and Punu, are genetically very similar to the other neighbouring male Bantu villagers, and overall, cannot be distinguished from them. A better understanding of this trait will certainly provide useful insights into the population history of sub-Saharan Africa.

Overall study of human genetic diversity of west-central Africa

West-central Africa is no longer a blank spot on the genetic map of Africa. The LCGB project allowed collecting DNA from some 1000 individuals, and currently 1400 well-documented samples of excellent quality are available. Recent studies based on these samples start yielding a good panorama of the present and the past of the populations of the region.

Many specific observations can be made. As an example only the case of the earlier mentioned Makina will be indicated here. Makina speak Shiwa and are currently being assimilated by the Fang. Our study reveals that this process, which started 2 or 3 centuries ago, has not yet affected the gene pool of the two populations in contact, whereas the cultural and linguistic impact of Fang on Shiwa is easily detected.

Future research

The interpretation of the emerging results raises the vital issue of a thorough collaboration with cultural anthropologists, archaeologists and other specialists. Mating patterns, descent systems, residence systems, war, (local) traditional slavery, social stratification, endogamy/exogamy, mobility, lifestyle, habitat, illness, food, group size, etc., are all factors that may account for the current genetic, linguistic and cultural diversity. In depth research on descent systems, marriage and residence strategies is currently carried out by cultural anthropologists working in Gabon and Lyon¹⁷. Archaeological research is ongoing (B. Clist, R. Oslistly). Additional cultural markers, such as artefacts, music, descent systems, technological skills, etc., will have to be taken into consideration in order to construct more complex multidimensional models. Multiplying the number of markers and diversifying their nature, will surely open up new perspectives and facilitate the search for relevant boundaries within these networks.

So far, sampling has privileged ethnic groups (our “populations”). These are certainly not the most relevant units for this type of investigation. The various ethnic groups have a strong social dimension, outweighing the biological component. The lineage, however, as a subsection of the ethnic group, seems to be a more meaningful object of investigation, this entity being much more biologically determined. Information about the lineages has been systematically included in our database and will reveal very useful for ongoing research intending to zoom in on specific issues.

New genetic (e.g. autosomal) markers, either microsatellites or sequencing of non-coding genome-wide, may shed new light on the complex relations between the groups examined here. Will these new markers corroborate the current findings or, on the contrary, reveal new correlations?

Linguists should now be focusing on the elaboration of more reliable regional classifications. Improvements can be achieved by using new computational and phylogenetic approaches (cf. Holden & Gray 2006, Alewijnse *et al.* 2007), by diversifying the linguistic data (i.a. morphological markers) and by taking into account new evidence from the study of special lexicon (flora, fauna, technology, toponymy, etc.) in relation with contact and population movements.

17. Research directed by Pr. Raymond Mayer (Laboratoire d'Anthropologie, Omar Bongo University, Libreville).

References

- Alewijnse, Bart, Nerbonne, John, Van der Veen, Lolke and Manni, Franz. 2007. "A computational analysis of Gabon varieties." In *Proceedings of the RANLP Workshop on Computational Phonology*, Petya Osenova (ed.). Recent Advances in Natural Language Phonology conference, Borovetz 2007. 3–12.
- Alves-Silva, Juliana, da Silva Santos, Magda, Guimarães, Pedro E. M., Ferreira, Allessandro C. S., Bandelt, Hans-Jürgen, Pena Sérgio D. J. and Prado, Vanio F. 2000. "The ancestry of Brazilian mtDNA lineages." *Am J Hum Genet* 67:444–6.
- Armour, John A., Anttinen, Tiiu, May, Celia A., Vega, Emilce E., Sajantila, Antti, Kidd, Judith R., Kidd, Kenneth K., Bertranpetit, Jaume, Pääbo, Svanto and Jeffreys, Alec J. 1996. "Minisatellite diversity supports a recent African origin for modern humans." *Nat Genet* 13(2):154–60.
- Bandelt, Hans-Jürgen and Forster, Peter. 1997. "The myth of bumpy hunter-gatherer mismatch distributions." *Am J Hum Genet* 61:980–83.
- Bandelt, Hans-Jürgen, Forster, Peter, Sykes, Bryan C. and Richards, Martin B. 1995. "Mitochondrial portraits of human populations using median networks." *Genetics* 141(2):743–53.
- Bandelt, Hans-Jürgen, Alves-Silva, Juliana, Guimarães, Pedro E., Santos, M.S., Brehm, António, Pereira, Luísa, Coppa, Alfredo, Larruga, José M., Rengo, Chiara, Scozzari, Rosaria, Torroni, Antonio, Prata, Maria J., Amorim, Antonio, Prado, Vanio F. and Pena, Sérgio D. J. 2001. "Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade." *Ann Hum Genet* 65:549–63.
- Bastin, Yvonne, Coupez, André and Mann, Michael. 1999. "Continuity and Divergence in the Bantu Languages: Perspectives from a Lexicostatistic Study." *Annales des sciences humaines* 162. Tervuren: Musée royal d'Afrique centrale.
- Batini, Chiara, Coia, Valentina, Battaggia, Cinzia, Rocha, Jorge, Pilkington, Maya M., Spedini, Gabriella, Comas, David, Destro-Bisol Giovanni and Calafell, Francesc. 2007. "Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of Central Africa." *Mol Phylogenet Evol.* 43(2):635–44.
- Bekombo-Priso, Manga and Laburthe-Tolra, Philippe (eds). 1981. "Le peuplement des régions centrale et méridionale." In *Contribution de la recherche ethnologique à l'histoire des civilisations du Cameroun*, Tardits C. (ed.). Colloques internationaux du CNRS. Paris, September 1973. Vol. 2: 579–86. Paris: Editions du CNRS.
- Beleza, Sandra, Gusmao, Leonor, Amorim, Antonio, Carracedo, Angel, Salas, Antonio. 2005. "The genetic legacy of western Bantu migrations." *Hum Genet* 117:366–75.
- Berniell-Lee, Gemma, Calafell, Francesc, Bosch, Elena, Heyer, Evelyn, Sica, Lucas, Mouguia-ma-Daouda, Patrick, van der Veen, Lolke, Hombert, Jean-Marie, Quintana-Murci, Lluís and Comas, David. 2009. "Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages" *Mol Biol Evol* (26: 1581–1589).
- Cann, Rebecca L., Stoneking, Mark and Wilson, Allan C. 1987. "Mitochondrial DNA and human evolution." *Nature* 325(6099):31–6.
- Cavalli-Sforza, Luigi L. 1986. *African Pygmies*. Orlando: Ed. Cavalli-Sforza L.L. Academic Press.
- Cavalli-Sforza, Luigi L. and Feldman Marcus W. 2003. "The application of molecular genetic approaches to the study of human evolution." *Nat Genet* 33 Suppl: 266–7.
- Chen, Yu S., Torroni, Antonio, Excoffier, Laurent, Santachiara-Benerecetti, Silvana and Wallace, Douglas C. 1995. "Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups." *Am J Hum Genet* 57:133–49.

- Clist, Bernard. 2005. *Des premiers villages aux premiers Européens autour de l'Estuaire du Gabon. Quatre millénaires d'interactions entre l'homme et son milieu*. PhD Thesis. Brussels: ULB.
- Coia, Valentina, Caglia, Alessandra, Arredi, Barbara, Donati, Francesco, Santos, Fabrício R., Pandya, Arpita, Taglioli, Luca, Paoli, Giorgio, Pascali, Vincenzo, Spedini, Gabriella, Destro-Bisol, Giovanni and Tyler-Smith, Chris. 2004. "Binary and microsatellite polymorphisms of the Y-chromosome in the Mbenzele pygmies from the Central African Republic." *Am J Hum Biol* 16: 57–67.
- Côrte-Real, Francisco, Carvalho, Mónica, Andrade, Lisa, Anjos, Maria J., Pestoni, Carmela, Lareu, Maria V., Carracedo, Angel, Vieira, Duarte N. and Vide, Maria C. 2000. "Chromosome Y STRs analysis and evolutionary aspects for Portuguese spoken countries." *Progress in Forensic Genetics* 8: 272–74.
- Cruciani, Fulvio, Santolamazza, Piero, Shen, Peidong, Macaulay, Vincent, Moral, Pedro, Olckers, Antonel, Modiano, David, Holmes, Susan, Destro-Bisol, Giovanni, Coia, Valentina, Wallace, Douglas C., Oefner, Peter J., Torroni, Antonio, Cavalli-Sforza, Luigi L., Scozzari, Rosaria and Underhill, Peter A. 2002. "A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes." *Am J Hum Genet* 70(5):1197–214.
- Destro-Bisol, Giovanni, Donati, Francesco, Coia, Valentina, Boschi, Ilaria, Verginelli, Fabio, Caglià, Alessandra, Tofanelli, Sergio, Spedini, Gabriella and Capelli, Christian. 2004. "Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors." *Mol Biol Evol* 21: 1673–82.
- Diop, Cheikh A. 1979 [1955]. *Nations nègres et culture. De l'antiquité nègre égyptienne aux problèmes culturels de l'Afrique Noire d'aujourd'hui*. Paris: Présence africaine.
- Diop, Cheikh A. 1981. *Civilisation ou barbarie? Anthropologie sans complaisance*. Paris: Présence africaine.
- Diop, Cheikh A. 1993 [1977]. *Antériorité des civilisations nègres: mythe ou vérité historique?* Paris/Dakar: Présence africaine.
- Guthrie, Malcolm. 1948. *The Classification of the Bantu Languages*. London: Oxford University Press for the International African Institute.
- Guthrie, Malcolm. 1967–1971. *Comparative Bantu*. Farnborough: Gregg International Publishers.
- Hammer, Michael F. 1995. "A recent common ancestry for human Y chromosomes." *Nature* 378(6555): 376–8.
- Hammer, Michael F., Karafet, Tatiana, Rasanayagam, Arani, Wood, Elizabeth T., Altheide, Tasha K., Jenkins, Trefor, Griffiths, Robert C., Templeton, Allan R. and Zegura, Stephen L. 1998. "Out of Africa and back again: nested cladistic analysis of human Y chromosome variation." *Mol Biol Evol* 15: 427–41.
- Hammer, Michael F., Karafet, Tatiana M., Redd, Alan J., Jarjanazi, Hamdi, Santachiara-Benerecetti Silvana, Soodyall, Himla and Zegura, Stephen L. 2001. "Hierarchical patterns of global human Y-chromosome diversity." *Mol Biol Evol* 18: 1189–203.
- Holden, Claire and Gray, Russell. 2006. "Exploring Bantu linguistic relationships using trees and networks." In *Phylogenetic Methods and the Prehistory of Languages*, Forster, Peter and Renfrew, Colin (eds). Cambridge: The MacDonald Institute for Archaeological Research. 19–31.
- Hombert, Jean-Marie, Medjo Mvé, Pither and Nguéma, Raymond. 1989. "Les Fang sont-ils Bantu?" *Pholia*, 4, Lyon: LAPHOLIA, CRLS, Université Lumière-Lyon 2. 133–47.
- Jorde, Lynn B., Watkins, W. Scott, Bamshad, Mike J., Dixon, Mary E., Ricker, Chris E., Seielstad, Mark T. and Batzer, Mark A. 2000. "The distribution of human genetic diversity: a com-

- parison of mitochondrial, autosomal, and Y-chromosome data." *Am J Hum Genet* 66(3):979–88.
- Kivisild, Toomas, Bamshad, Michael J., Kaldma, Katrin, Metspalu, Mait, Metspalu, Ene, Reidla, Maere, Laos S., Parik, Jüry, Watkins, W. Scott, Dixon, Mary E., Papiha, Surinder S., Mastana, Sarabjit S., Mir M.R., Ferak, Vladimir and Vilems, Richard. 1999. "Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages." *Curr Biol* 9(22):1331–4.
- Knight, Alec, Underhill, Peter A., Mortensen, Holly M., Zhivotovsky, Lev A., Lin, Alice A., Henn, Brenna M., Louis, Dorothy, Ruhlen, Mehrit and Mountain, Joanna L. 2003. "African Y chromosome and mtDNA divergence provides insight into the history of click languages." *Curr Biol* 13:464–73.
- Luis, Javier R., Rowold, Diane J., Regueiro, Maria, Herrera, Rene J. 2004. "The Levant versus the Horn of Africa: Evidence for Bidirectional Corridors of Human Migrations." *Am J Hum Genet* 74:532–44.
- Macaulay, Vincent, Hill, Catherine, Achilli, Alessandro, Rengo, Chiara, Clarke, Douglas, Meehan, William, Blackburn, James, Semino, Ornella, Scozzari, Rosaria, Cruciani, Fulvio, Taha, Adi, Shaari, Norazila K., Raja, Joseph M., Ismail, Patimah, Zainuddin, Zaffarina, Goodwin, William, Bulbeck, David, Bandelt Hans-Jürgen, Oppenheimer, Stephen, Torroni, Antonio and Richards, Martin 2005. "Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes." *Science* 308(5724):1034–6.
- Maley, Jean. 2001. "La destruction catastrophique des forêts d'Afrique centrale survenue il y a environ 2500 ans exerce encore une influence majeure sur la répartition actuelle des formations végétales." *Systematic & Geography of Plants*, 71: 777–796.
- Mateu, Eva, Comas, David, Calafell, Francesc, Pérez-Lezaun, Anna, Abade, Augusto and Bertranpetit, Jaume. 1997. "A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and Sao Tome, Gulf of Guinea." *Ann Hum Genet* 61 (Pt 6): 507–18.
- Mayer, Raymond. 1987. "Langues des groupes pygmées du Gabon: un état des lieux." *Pholia*, 2. Lyon: LAPHOLIA, CRLS, Université Lumière-Lyon 2. 111–24.
- Mba-Nkoghe, Jules. 2001. *Description linguistique du fang du gabon, parler atsi: phonologie, morphologie, syntaxe, lexique*. PhD thesis. Paris: Paris 3 (Sorbonne).
- Medjo Mvé, Pither. 1997. *Essai sur la phonologie panchronique des parlers fang du Gabon et ses implications historiques*. PhD Thesis. Lyon: Université Lumière-Lyon 2.
- Mouguiama-Daouda, Patrick. 2005. *Contribution de la linguistique à l'histoire des peuples du Gabon. La méthode comparative et son application au bantu*. Collection Sciences du Langage. Paris: CNRS Editions.
- Mouguiama-Daouda, Patrick and Van der Veen, Lolke J. 2005. "B10-B30: conglomerat phylogénétique ou produit d'une hybridation". In *Studies in African Comparative Linguistics, with special focus on Bantu and Mande*, Bostoen, Koen and Maniacky, Jacky (eds). Tervuren: Royal Museum for Central Africa (RMCA/MRAC), Sciences Humaines, 1781–857.
- Obenga, Théophile. 1985. *Les Bantu. Langues. Peuples. Civilisations*. Paris: Présence africaine.
- Obenga Théophile. 1993. *Origine commune de l'égyptien ancien, du copte et des langues négro-africaines modernes. Introduction à la linguistique historique africaine*. Paris: L'Harmattan.
- Patin, Etienne, Laval, Guillaume, Barreiro, Luis B., Salas, Antonio, Semino, Ornella, Santachiara-Benerecetti, Silvana, Kidd, Kenneth, Kidd, Judith, van der Veen, Lolke, Hombert, Jean-Marie, Gessain, Antoine, Froment, Alain, Bahuchet, Serge, Heyer, Evelyne and Quintana-Murci, Lluís. 2009. "Inferring the Demographic History of African Farmers and Pygmy Hunter-Gatherers Using a Multilocus Resequencing Data Set" *PLoS Genet* 5: e1000448.

- Pereira, Luísa, Gusmao, Leonor, Alves, Cíntia, Amorim, Antonio and Prata, Maria J. 2002. "Bantu and European Y-lineages in Sub-Saharan Africa." *Ann Hum Genet* 66: 369–78.
- Plaza, Susana, Calafell, Francesc, Helal, Ahmed, Bouzerna, Noureddine, Lefranc, Gérard, Bertranpetit, Jaume and Comas, David. 2003. "Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean." *Ann Hum Genet* 67:312–28.
- Plaza, Susana, Salas, Antonio, Calafell, Francesc and Comas, David. 2004. "Insights into the Western Bantu dispersal: mtDNA lineage analysis in Angola." *Hum Genet* 115:439–47.
- Quintana-Murci, Lluís, Quach, Hélène, Harmant, Christine, Luca, Francesca, Massonnet, Blaindine, Patin, Etienne, Sica, Lucas, Mouguiama-Daouda, Patrick, Comas, David, Tzur, Shay, Balanovsky, Oleg, Kidd, Kenneth K., Kidd, Judith R., Van der Veen, Lolke, Hombert, Jean-Marie, Gessain, Antoine, Verdu, Paul, Froment, Alain, Bahuchet, Serge, Heyer, Evelyne, Dausset, Jean, Salas, Antonio and Behar, Doron M. 2008. "Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers." *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 105:5, 1596–601.
- Quintana-Murci, Lluís, Semino, Ornella, Bandelt, Hans-Jürgen, Passarino, Giuseppe, McElreavey, Ken and Santachiara-Benerecetti A. Svilvana. 1999. "Genetic evidence of an early exit of Homo sapiens from Africa through eastern Africa." *Nat Genet* 23:437–41.
- Rando, Juan C., Pinto, Francisco, González, Ana M., Hernández, Marta, Larruga, José M., Cabrera, Vicente M., Bandelt Hans-Jürgen. 1998. "Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations." *Ann Hum Genet* 62:531–50.
- Rexová, Kateřina, Bastin, Yvonne and Frynta, Daniel. 2006. "Cladistic analysis of Bantu languages: a new tree based on combined lexical and grammatical data." *Naturwissenschaften* 93: 189–94.
- Salas, Antonio, Richards, Martin, De la Fé, T., Lareu, Maria V., Sobrino, Blanca, Sánchez-Diz, Paula, Macaulay, Vincent and Carracedo, Angel. 2002. "The making of the African mtDNA landscape." *Am J Hum Genet* 71:1082–111.
- Salas, Antonio, Richards, Martin, Lareu, Maria V., Scozzari Rosaria, Coppa Alfredo, Torrioni, Antonio, Macaulay, Vincent and Carracedo, Angel. 2004. "The African diaspora: mitochondrial DNA and the Atlantic slave trade." *Am J Hum Genet* 74:454–65.
- Scozzari, Rosaria, Cruciani, Fulvio, Malaspina, Patrizia, Santolamazza, Piero, Ciminelli, Bianca M., Torrioni, Antonio, Modiano, David, Wallace Douglas C., Kidd, Kenneth K., Olckers, Antonel, Moral, Pedro, Terrenato, Luciano, Akar, Nejat, Qamar, Raheel, Mansoor, Ahmed, Mehdi, S. Qasim, Meloni, Gian Franco, Vona, Giuseppe, Cole, D.E., Cai, W. and Novelletto, Andrea. 1997. "Differential structuring of human populations for homologous X and Y microsatellite loci." *Am J Hum Genet* 61(3):719–33.
- Scozzari, Rosaria, Cruciani, Fulvio, Santolamazza, Piero, Malaspina, Patrizia, Torrioni, Antonio, Sellito, Daniele, Arredi Barbara, Destro-Bisol, Giovanni, De Stefano, Gian Franco, Rickards, Olga, Martínez-Labarga, Cristina, Modiano, David, Biondi, Gianfranco, Moral, Pedro, Olckers, Antonel, Wallace, Douglas C. and Novelletto, Andrea. 1999. "Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations." *Am J Hum Genet* 65:829–46.
- Soodyall, Himla, Vigilant, Linda, Hill, Adrian V. S., Stoneking, Mark and Jenkins, Trefor. 1996. "MtDNA control-region sequence variation suggests multiple independent origins of an Asian-specific 9-bp deletion in sub-Saharan Africans." *Am J Hum Genet* 58:595–608.

- Stringer, Christopher B. and Andrews, P. 1988. "Genetic and fossil evidence for the origin of modern humans." *Science* 239(4845):1263–8.
- Swisher, Carl C. III, Rink, W. Jack, Anton, S. C., Schwarcz, Henry P., Curtis, Garniss H., Suprijo, A. and Widiastomo. 1996. "Latest *Homo erectus* of Java: potential contemporaneity with *Homo sapiens* in southeast Asia." *Science* 274(5294):1870–4.
- Thomas, Mark G., Parfitt, Tudor, Weiss, Deborah A., Skorecki, Karl, Wilson, James F., le Roux, Magdel, Bradman, Neil and Goldstein, David B. 2000. "Y chromosomes traveling south: the cohen modal haplotype and the origins of the Lemba - the "Black Jews of Southern Africa". *Am J Hum Genet* 66:674–86.
- Tishkoff, Sarah A., and Kidd, Kenneth K. 2004. "Implications of biogeography of human populations for 'race' and medicine." *Nat Genet* 36:S21–7.
- Tishkoff, Sarah A., Dietzsch, Erin, Speed, William, Pakstis, Andrew J., Kidd, Judith R., Cheung, Kei-Hoi, Bonne-Tamir, Batsheva, Santachiara-Benerecetti, A. Silvana, Moral, Pedro, Krings, Matthias, Pääbo, Svanto, Watson, Elizabeth, Risch, N., Jenkins, Trefor and Kidd, Kenneth K. 1996. "Global patterns of linkage disequilibrium at the CD4 locus and modern human origins." *Science* 271(5254):1380–7.
- Torrioni, Antonio, Rengo, Chiara, Guida, Valentina, Cruciani, Fulvio, Sellitto, Daniele, Coppa, Alfredo, Calderon, Fernando L., Simionati, Barbara, Valle, Giorgio, Richards, Martin, Macaulay, Vincent and Scozzari, Rosaria. 2001. "Do the four clades of the mtDNA haplogroup L2 evolve at different rates?" *Am J Hum Genet* 69:1348–135.
- Trilles, Reverend P. 1912a. *Chez les Fang ou quinze années de séjour au Congo français*. Paris: Société Augustin-Desclée de Brouwer.
- Trilles, Reverend P. 1912b. Le totémisme chez les Fang. *Anthropos*.
- Trilles, Reverend P. 1931. *Mille lieues dans l'inconnu. En pleine forêt équatoriale, chez les Fang*. Paris: Desclée de Brouwer.
- Trovoada, M. J., Alves, Cíntia, Gusmao, Leonor, Abade, Augusto, Amorim, Antonio and Prata, Maria J. 2001. "Evidence for population sub-structuring in São Tomé e Príncipe as inferred from Y-chromosome STR analysis." *Ann of Human Genetics* 65:271–83.
- Underhill, Peter A., Passarino, Giuseppe, Lin, Alice A., Shen, Peidong, Mirazon Lahr, Marta, Foley, Robert A., Oefner, Peter J. and Cavalli-Sforza, Luigi L. 2001. "The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations." *Ann Hum Genet* 65:43–62.
- Van der Veen, Lolke J. 2006. Gabon: Language situation. In *Encyclopedia of Language and Linguistics (ELL2)*, Brown K. (ed.). Second Edition. Oxford, Elsevier Ltd.
- Wainscoat, James S., Hill, Adrian V. S., Boyce, A. L., Flint, J., Hernandez, Marta, Thein, Sweet L., Old, John M., Lynch, J.R., Falusi, Adeyinka G., Weatherall, David J. *et al.* 1986. "Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms." *Nature* 319(6053):491–3.
- Watson, Elizabeth, Forster, Peter, Richards, Martin, Bandelt Hans-Jürgen. 1997. "Mitochondrial footprints of human expansions in Africa." *Am J Hum Genet* 61:691–704.
- Wood, Elizabeth T., Stover, Daryn A., Ehret, Christopher., Destro-Bisol, Giovanni, Spedini, Gabriella, McLeod, Howard, Louie, Leslie, Bamshad, Mike, Strassmann, Beverly I., Soodyall, Himla and Hammer, Michael F. 2005. "Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes." *Eur J Hum Genet* 13: 867–76.
- Y-Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339–48.

The Berber and the Berbers

Genetic and linguistic diversities

Jean-Michel Dugoujon¹, Clotilde Coudray¹, Antonio Torroni², Fulvio Cruciani³, Rosaria Scozzari³, Pedro Moral⁴, Naima Louali^{†5} and Maarten Kossmann⁶

In Northern Africa, two main blocks of languages are spoken: the vernacular Arabic varieties and the Berber languages. Our objective is to correlate linguistic characters to genetic data. The precise language affiliation of the samples is controlled. Relationships with Arabic speakers are examined. Four informative genetic markers have been studied: two are autosomal (Gm allotypes and Alu sequences) and the other two are of maternal (mitochondrial DNA) and paternal (Y chromosome) lineages. Four populations were studied (514 samples): three are from Morocco (Asni, Bouhria and Figuig oasis) and the fourth from Egypt (the Siwa oasis). The results were compared with previously published analyses. The genetic differentiation between North-Western and North-Eastern African Berbers (Siwan) seems to be correlated with linguistic studies.

The Berber case

Today Berbers constitute a wide set of different human groups inhabiting North Africa. Even though the question of their origin is an intriguing and controversial subject, the presence of Berbers is often assumed to go back at least to the Capsian era (10,000–4,700 years ago), even though one should be aware of the fact that this industry derives from older cultures, and that the definition of “Berbers” during

-
1. Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse (AMIS), CNRS-Université Paul Sabatier, Toulouse 3, Toulouse, France
 2. Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia, Italy
 3. Dipartimento di Genetica e Biologia Molecolare, Sapienza Università di Roma, Rome, Italy
 4. Departamento de Biología Animal-Antropología, Universidad de Barcelona, Spain
 5. Laboratoire Dynamique du Langage, CNRS-Université Lumière Lyon 2, Lyon, France
 6. Department of Languages and Cultures of Africa, Leiden University, The Netherlands

this period is an awkward problem (see below). Nevertheless, it was only during the Neolithic transition (around 6000 years ago in the Saharan areas and 5000 years ago in the Maghreb) that Northern Africa was incontestably marked by various cultural events. Northern Africa experienced a long and complicated history with many invasions, conquests and migrations by, among others, Phoenicians, Romans, Vandals and Byzantines (Brett and Fentress 1996). The most significant event was the Arab conquest, which started in the 7th century, when North-Africans were converted to Islam and Arabic became the unique official language employed. In spite of strong resistance, Berbers acquiesced to Islamic authority. Refractory groups were driven out and constrained to more isolated areas. This troubled past directly influenced the geographical distribution of Berber communities which are nowadays scattered in a vast region extending from Mauritania to Egypt (the Siwa oasis) and from south of the Sahara desert to the mountainous areas bordering the Mediterranean.

At this point, a short note of terminology is due. The term “Berbers” is used in different contexts in different ways. In a most general way, “Berbers” is a cover term for all non-Semitic autochthonous populations in Northern Africa. This definition does not in itself have any linguistic or genetic implications: “autochthonous” populations may be very diverse in their genetic, linguistic, or cultural affiliations. Of course, by choosing the term Berber, a certain continuity with modern populations speaking Berber languages is suggested. As autochthonous is a vague term in itself, Berber-ness is sometimes extended to all or almost all people living in Northern Africa: one may speak about “Arabicized Berbers”, which is best interpreted as genetically autochthonous but linguistically (and maybe culturally) Arabic populations. In a more delimiting fashion, “Berber” is sometimes used to denote certain cultural phenomena; thus Malika Hachid (2000) sees a long cultural continuity between certain prehistoric populations and modern Berbers because of certain shared cultural characteristics. This makes her speak of these prehistoric populations as “Berbers”. Similarly, it is often claimed (and not without reason) that rural Northern Africa culture is relatively homogeneous on a local level, and that the difference between “Arabs” and “Berbers” is of little importance here. Again, often this culture is referred to as “Berber” culture. In all these definitions, the difference between “Berbers” and “Arabs” is extremely fluid, and one has the feeling that they are more linked to certain stereotypes (the decadent Arabic town-dweller as opposed to the freedom-loving Berber peasant in some recent Berber nationalistic discourses) than to a clear view of the North-African reality.

The only element which seems to delimit “Arabs” and “Berbers” in a reasonably discrete way is language. In Northern Africa, two main blocks of languages are spoken. On the one hand, there are the vernacular Arabic varieties, on the other hand there are the Berber languages. Arabic and Berber are ultimately

related, as Semitic (to which Arabic belongs) and Berber form coordinate branches in the Afro-Asiatic linguistic phylum, but this affiliation has no impact on the recognition of the two blocks. It is relatively easy to define a “Berber” as someone who has a Berber language as his or her mother tongue, while “Arabs” would be those people who have been raised in (vernacular) Arabic from their early childhood on. Even this definition is not without problems (how to categorize perfect bilinguals, or is an Arabic-speaking child of two Berber-speaking parents an “Arab”), but it works quite well.

It may be clear, under these circumstances, that a purely linguistic definition is by no means a logical correlate to a populations-genetic unity. The case of Arabic-speaking populations who carry Berber tribal names (such as the Branes in the Moroccan Jbala) is illustrative: there is no doubt that many Arabic-speaking populations acquired Arabic through language shift somewhere in their history, so that there is no expected correlation between speaking Arabic and having a certain genetic profile. The problem is somewhat softened by the fact that, as far as we know, during the history of Northern Africa, Arabic-speaking populations never shifted language on a larger scale to Berber, so that those populations who nowadays speak a Berber language may reasonably be considered to form a (cultural) continuation of pre-Arabic autochthonous populations.

This, however, does not at all imply that there is no admixture of “Arabs” in Berber-speaking populations. Northern-African culture favours endogamy, so most people will marry people from their own group. On the other hand, there exist various ways in which people may enter a group. For example, Northern African culture has a long tradition of “adoption”, i.e. a person or a whole group of persons (a family, a subgroup of a tribe) may be adopted by a group. After this adoption, the new persons become part of the larger entity. They are acculturated and take over the language of the group they entered. In this way, many speakers of Arabic will have become part of Berber speaking groups and shifted to Berber.

A second problem in the correlation of linguistic facts comes from the history of the Berber languages. Northern Africa has been populated for a very long time, and there seem to be few major population breaks since Mesolithic times. Berber languages, on the other hand, constitute a fairly homogeneous language group (many scholars, e.g. Chaker 1995: 9, even prefer to speak of a single Berber language). As linguistic diversity in a language group tends to increase with time, one may assume that the earliest ancestor of the Berber languages is not that far away in time. Several scholars have suggested that the level of diversity inside Berber is similar to that inside the Germanic or the Romance language groups. If diversification and time depth were to correlate in the same way in these European language groups as in Berber, this would imply a time depth of about 2000–2500 years only (Louali and Philippson 2004). Blench (2001) has suggested a different

scenario, which is based on the idea that highly mobile nomadic populations in open terrain may maintain much more linguistic homogeneity because of the constant encounters and mixing of groups. Whatever the merits of the model as such, it cannot apply to most of the actual habitats of Berber speaking groups, which are mountain areas.

If one adopts a dating for the proto-Berber language somewhere in the first millennium BC, one faces two important facts. In the first place, most of Northern Africa was well-populated at that time. In the second place, there are no archaeological traces of a major demic expansion during this period in this region. As proto-languages (or their immediate ancestor) are by definition homogeneous, and therefore doubtlessly restricted to a smaller geographic domain, one must assume that the language started to spread from a certain point (presumably in Northern Africa). In order to understand the absence of traces of major demic changes during the period of the linguistic expansion (and well beyond that period), one has to assume that it was a cultural, and not a demic phenomenon. That is to say, Berber expanded over Northern Africa, not because “Berbers” spread over the continent, but because populations shifted from their original language to Berber.

This implies *par force* that there were, during the first millennium BC, populations living in Northern Africa, who spoke a language different from the ancestor of the modern Berber languages. Louali and Philippson (2004) have suggested that these may have been sister languages of Proto-Berber, i.e. languages belonging to the Berber branch of Afroasiatic, but which are not the immediate ancestors of modern Berber. This is very possible, but there are only few (if any) positive indications for this. It is equally possible to assume languages which were not at all related to Berber. Moreover, there is no reason to posit linguistic homogeneity in Northern Africa before the Berber linguistic expansion – there may have been dozens of languages, belonging to different language families.

In our “Berber” project, the objectives are to correlate linguistic and historical characters to genetic data studies. We are in a uniquely favourable position to describe the linguistic and genetic diversities of many North African Berber populations because the precise language affiliation of the donors is controlled, enhancing the reliability of the genetic results. So, through the study of various polymorphic markers, we can provide a detailed view of the Berber genetic pool and examine the relationships not only between Berber-speaking populations, but also between Berbers and Arabic-speakers. Our assumption is that, in the course of time, the various populations which migrated to North Africa, in addition to their demographic impact, have probably left a footprint in the Berber genetic pool.

Genetic results

To describe the present-day genetic diversity of Berber populations and to relate it to the multiple past migrations which took place in North Africa, we present here data on four informative genetic markers. Two are autosomal markers: the GM allotypes and the Alu sequences. The other two are markers of maternal (mitochondrial DNA) and paternal (Y chromosome) lineages.

Where possible, results are illustrated by the genetic data of some populations we sampled. Three of them are Berber speakers from Morocco (Asni, Bouhria and Figuig) and the fourth is the Berber population from Egypt (the Siwa oasis). The inclusion in the study of a sample of Berbers from North-Eastern Africa (Siwa Berbers) is particularly relevant, since only Berbers from North-West Africa have been genetically studied so far.

Asni is located in the Tacheddirt valley in the High Atlas Mountains (47 km from Marrakech). The Berbers from Asni speak Tashelhiyt Berber and belong to the Rhiraya tribe. Sidi Bouhria is located in North-Eastern Morocco (Oujda wilaya). Its population belongs to the Beni Moussi Roua fraction of the Beni Iznasen tribe. Figuig is located in the extreme east of Morocco, at the Algerian border, at the junction of the High Plateaus and the North-Western edge of the Sahara. It is a pre-Saharan oasis comprising seven villages, one in the lower part of the oasis, Zenaga, the other six in the higher part (also called High Figuig). Linguistically, there is a major break between the oasis dwellers, who are predominantly Berber speakers (for a description of Figuig Berber, see Kossmann 1997), and the nomads roaming the mountains and the steppes surrounding the oasis, who are all speakers of Arabic. In addition, there are a few groups who are reported to speak Arabic inside the Figuig oasis, who mainly live in some of the villages in High Figuig. The Berber population we collected lives in the largest Figuig village, Zenaga. The Siwa oasis is located in the western Egyptian desert, at 300 km from the Mediterranean coast and 25 km from the Libyan border. The Berber language of Siwa (on which see Laoust 1931; Louali and Philippson 2005; Vycichl 2005) poses intricate problems from the point of view of a historical linguist. On the one hand, in many respects Siwan is quite similar to the Berber languages in central Libya (El Foqaha, Sokna, Djebel Nefusa) and even more to the west (cf. Kossmann 1999: 31). At this point it differs from its immediate neighbour to the west, Awdjilah Berber, which diverges from all other Berber languages of the region, and which may constitute an early branching of the Berber language family (cf. Naumann 1999). On the other hand, Siwan has a number of features which are either unique in Berber, or which are only encountered in Siwa and in Awdjilah.

GM allotypes

Four allotypic systems have been described according to immunoglobulin classes or chain types. Among these, the most useful system in characterizing human groups is investigated: the GM system. GM antigenic determinants are present on the heavy chains of 3 of the 4 subclasses of IgG (IgG1, IgG2 and IgG3) (Lefranc and Lefranc 1990). GM allotypes are encoded by closely linked alleles (on chromosome 14) that are inherited in fixed combinations (rare recombinations) called haplotypes. The qualitative and quantitative differences between GM haplotypes are an excellent means for comparing human populations from different geographic areas (genetic structure, degree of admixture, gene flow, migration and evolutionary history) (Cavalli-Sforza *et al.* 1994; Dugoujon *et al.* 2004; Steinberg and Cook 1981).

Various works on GM polymorphism of North-Africans have been published. They relate to Arabic-speaking and Berber-speaking populations from Tunisia (Chaabani *et al.* 1984; Chaabani and Cox 1988; Fadhlouli-Zid *et al.* 2004a; Lefranc *et al.* 1979; Loveslati *et al.* 2001), Morocco (Coudray *et al.* 2006; Dugoujon *et al.* 2005), and Algeria (Dugoujon *et al.* 2004; Lefèvre-Witier 1982). Results showed that Berber populations had an allotypic structure composed of more than 80 per cent of haplotypes frequently met in Europeans, and of from 10 to 20 per cent of haplotypes common in sub-Saharan populations. This GM haplotype distribution is reflected by a genetic proximity between North-African and South-European populations. Results also showed that Berber-speakers and Arabic-speakers are not genetically differentiated. However, a striking result is observed in Berber oases: in Siwa (Coudray *et al.* 2008) and Figuig (Dugoujon, unpublished results) the sub-Saharan haplotypes have a high frequency.

In our study, we analyzed GM haplotype frequencies of 15 Berber and 4 Arab populations (Fig. 1). While the Berber GM pool is broadly of the same composition as previously published, the PCA shows a very important genetic difference of Egyptian Berbers from Siwa, as compared to all other North-African Berber-speaking or Arabic-speaking populations. This population situation is explained by great differences in the GM haplotype distribution. We observe a distinction between the haplotypes frequently met in European populations (H1, H2, H3, and H4), one of the haplotypes commonly present in Asian populations (H5) and the GM haplotypes present at high rates in sub-Saharan groups (H6, H7, H8, H9, and H10). The first are very frequent in the Maghreb (from 66% to 92%), the second is observed at around 1 per cent in Ghardaïa, Lower Kabylia and Siwa, and finally the last are present at 49 per cent in Siwa, and have a lower frequency in the Maghreb (from 8% to 34%). The graph also shows that there is a genetic proximity between Berber-speakers and Arabic-speakers. An Analysis of Variance (ANOVA) statistically confirmed the result. Finally, results show that in Figuig, the Berber popula-

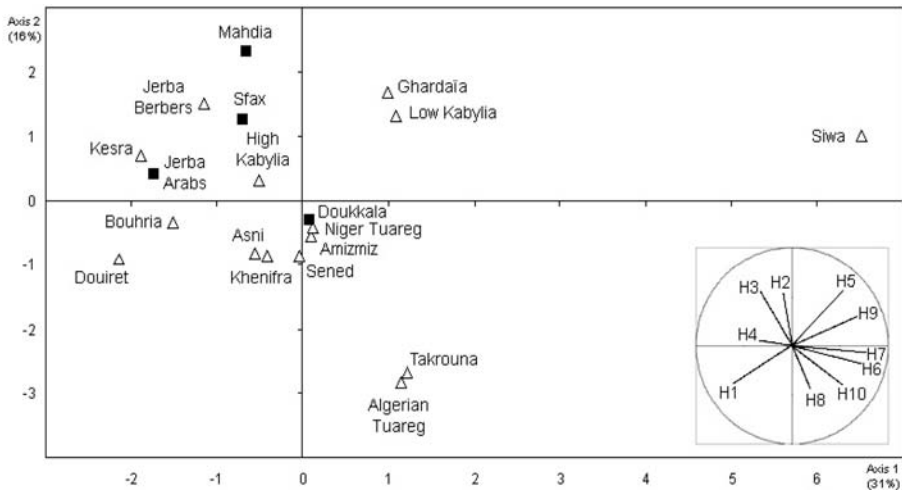


Figure 1. Principal Component Analysis (PCA) of GM haplotype frequencies in 19 Berber and Arab populations. The PCA was drawn on the basis of GM haplotype frequencies of 15 Berber (Δ) and 4 Arab ($\blacksquare$) North African populations. Variables (GM haplotypes) explaining the population distribution along the axes 1 and 2 are also shown: H1: Gm21,28;1,17;... ; H2: Gm21,28;1,2,17;... ; H3: Gm5*;3;23, H4: Gm5*;3;... ; H5: Gm10,11,13,15,16;1,17;... ; H6: Gm5*;1,17;... ; H7: Gm5*,28;1,17;... ; H8: Gm10,11,13,15, $\pm$ 28;1,17;... ; H9: Gm5,6,10,11,14, $\pm$ 28;1,17;... ; H10: Gm5,6,11,24, $\pm$ 28;1,17;... 5* stands for 5, 10, 11, 13, 14

tions present the highest frequencies of GM haplotypes commonly present in sub-Saharan populations (more than 50%).

Alu sequences

Other autosomal markers are biallelic and multiallelic polymorphisms such as SNPs and Alu insertions (18 on autosomes and 4 on the X chromosome) with low mutation rates, and several STRs with higher mutation rates (see, for instance, Athanasiadis *et al.* 2007; Esteban *et al.* 2004, 2005, 2006; Gonzalez-Perez *et al.* 2007). The use of markers with different mutational rates can allow us to identify demographic processes which occurred at different times in the past. In fact, the polymorphic Alu insertions display a very remarkable genetic stability and a clear polarity of the evolutionary change (with defined ancestral and derived alleles), and so they work very well for the reconstruction of ancestral population relationships. Besides, the combined analysis of some STRs linked to some Alu loci allows the definition of specific allele combinations (haplotypes) as well as the analysis of the STR variation in

the ancestral and derived Alu alleles that can be effective in considering detailed population relationships and calculate times of divergence between human groups.

Our results concerning the genetic distances and allele frequency variances indicate an important differentiation within the Berber populations. A detailed analysis of the distribution of the genetic distances indicates that the Egyptian Siwa population is clearly the most differentiated Berber population according to the variation of the different polymorphisms (autosomal and X-chromosome Alus, autosomal and X-chromosome STRs), followed by Mozabites, South-Moroccan Berbers, and Moroccan Arab-speakers present a peripheral position only in some analyses. In fact, the main genetic boundary in the Mediterranean basin separates Siwa Berbers from the rest.

Looking at potential factors for this within Berber genetic differentiation, we found particular allele STR/Alu combinations (haplotypes of both the CD4 and DM loci with a link to the Alu derived allele) that seem to be specific to North-African populations. So, the 110(-) haplotype of the CD4 locus is present among Berber populations with frequencies up to 7 per cent; it is shared by only two Southern European groups but it is absent in Sub-Saharan Africa and Central Europe. On the other hand, the DM 107(-) combination is present in all Berber groups (1.5 to 5% frequencies) but only shared, at lower frequencies, with three Southern European populations. The age of these two haplotype combinations has been estimated as 36,000 YBP for the CD4 110(-) and 48,000 YBP for the DM 107(-) combination.

The information provided by Alu/STR haplotype distribution probably explains only part of the genetic differentiation among Berbers. An analysis of the Sub-Saharan African gene flow can help to understand other genetic factors underlying the current differentiation between Berber populations. The results from the Alu data indicate a south-to-north decreasing gene flow with values of 6 per cent in Moroccan Arabs, North East Atlas and High Atlas Asni Berbers, intermediate values in Mozabites (9%) and Middle Atlas (12%), and the highest estimate in High Atlas Amizmiz Berbers (17%). The Siwa Berbers are the exception because the Sub-Saharan African genetic contribution in this sample is negligible. When we take into account the results provided by the Alu/STR combination, admixture proportions among North Africans are much more homogeneous. North East Atlas Berbers exhibit 16 per cent of Sub-Saharan African admixture. The remaining samples showed values close to 35 per cent and the Siwa Berbers are again the exception. Surprisingly, this time they show the highest admixture value (51%).

Mitochondrial DNA

Since the early 1980s, mitochondrial DNA (mtDNA) has been widely used for human population studies, because of its high copy number (hundreds to thousands of copies per cell), its maternal inheritance, and its evolutionary rate greater than those of the

nuclear DNA. MtDNA polymorphism has been generated by a sequential accumulation of mutations along radiating maternal lineages, and a mitochondrial haplogroup tree has been defined (see a simplified version in Figure 2). A detailed analysis of human mtDNA is a powerful tool in reconstructing the maternal population history of a region, such as Northern Africa, which is at the crossroads of Europe, the Middle East, and sub-Saharan Africa, regions for which mtDNA phylogeography is known in detail.

MtDNA analyses had previously been made of North African populations (Achilli *et al.* 2004; 2005; Brakez *et al.* 2001; Cherni *et al.* 2005; Côté-Real *et al.* 1996; Coudray 2006; Ennafaa *et al.* 2009; Fadhlouli-Zid *et al.* 2004b; Kéfi *et al.* 2005; Krings *et al.* 1999; Loveslati *et al.* 2006; Macaulay *et al.* 1999; Olivieri *et al.* 2006; Plaza *et al.* 2003; Rando *et al.* 1998; Stevanovitch *et al.* 2003). Authors show that Berbers have a mitochondrial genetic structure comprising lineages that are also frequently observed in Europe (H, V, J, T), as well as haplogroups present at high rates in Northern Asia (M) and the sub-Sahara (L2, L3). The U6 haplogroup is detected at relatively high frequencies in the Berbers from the Maghreb, and up to 30 per cent in Algerian Mozabite.

For the populations we considered, the Berber mitochondrial pool was observed to be composed of 44 per cent (in Figuig) to 87 per cent (in Jerba) of H, HV, J, T, and U* haplogroups (frequent in West Eurasian populations), of 13 per cent (in Bouhria) to 47 per cent (in Figuig) of L sub-Saharan haplogroups, of 0 to 16 per cent (in Siwa) of M1 haplogroups (frequent in East-Africa) and of 0 to 28 per cent (in Ghardaia)

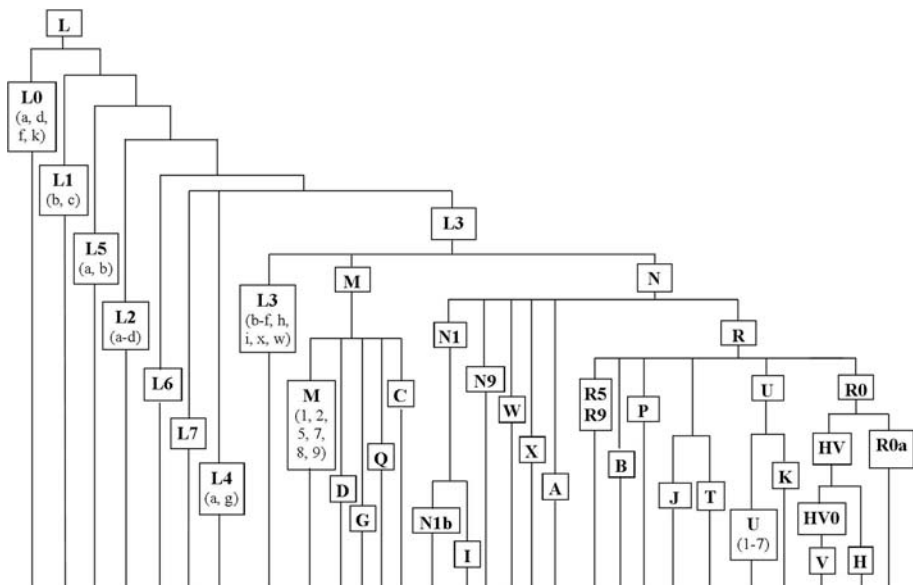


Figure 2. A simplified version of the mitochondrial DNA haplogroup tree. For details, see Coudray 2006.

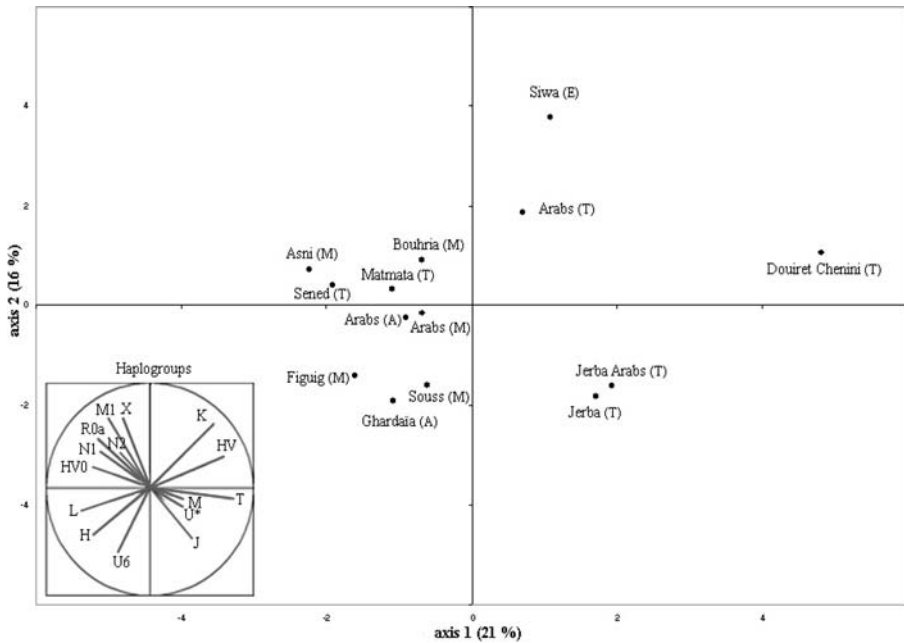


Figure 3. Principal Component Analysis (PCA) of mtDNA haplogroup profiles of 14 Berber and Arab populations.

The “A”, “E”, “M” and “T” letters refer to the country where each population lives: Algeria, Egypt, Morocco and Tunisia, respectively. Haplogroups and subhaplogroups included in the PCA are as follows: H (including H* and all H subhaplogroups), HV0 (including HV0* and V), HV (including HV* and HV1), R0a, J, T, U* (including all U lineages except U6 and K), U6, K, N1 (including N1 and I lineages), N2 (represented mainly by W), X, M (including M* and all M subhaplogroups except M1), M1, and L (including L0, L1, L2, L3, L4 and L5).

of the predominantly North-African U6 haplogroup. When we compared mitochondrial haplogroup frequencies between North-African Berber and Arabic groups (Fig. 3.), the PCA first reveals a significant genetic differentiation between Berbers from the Maghreb and Egyptian Berbers. Five haplogroup families underlined this differentiation. First, H, which is by far the most frequent lineage in Western European populations, accounts for 24 to 38 per cent of the Moroccan gene pool whereas only one individual from Siwa bears this lineage. Then U5, which is the most ancient haplogroup of the U* family originating in the Near East, accounts for 16 per cent of Siwi lineages whereas only one Berber from Asni also bears this sequence. Even if L African lineages, taken as a whole, are observed at slightly similar frequencies in Morocco (14% to 47%) and in Siwa (20%), their qualitative distribution is different from one area to another. For example, L1b and

L3e5 subclades, mainly detected in Western and Central African populations, are more frequent in the Maghreb (7% to 10% and 3% to 17%) than in Siwa (1%). Conversely, the L3e1 South-East African clade and the L0a1, L3i2, L4* and L4b2 East-African lineages are only observed in the Egyptian Berbers. The East-African M1 clade has a frequency four times higher in Siwa (16%) than in Morocco (from 2.6% to 4.2%). Finally, the North African U6 lineage was only observed in the Moroccan samples. As seen above for other markers, the PCA therefore shows that North African Berbers and Arabs are not genetically differentiated (result confirmed by an ANOVA) (see the synthesis in Coudray *et al.*, 2009).

Y chromosome

The male-specific portion of the human Y chromosome (MSY) is paternally transmitted and haploid. Patrilinearity and haploidy translate into increased levels of population subdivision compared to the autosomes, and the lack of recombination permits the reconstruction of an unequivocal MSY phylogeny which can be related to the geographic distribution of the Y haplogroups in an approach known as phylogeography (Avice *et al.* 1987; Underhill *et al.* 2001). These features make the DNA sequence variation of the MSY an invaluable tool for the study of modern human evolution.

To date, only a limited number of Berber-speaking populations from northern Africa have been studied with respect to biallelic MSY diversity (Arredi *et al.* 2004; Bosch *et al.* 2001; Cruciani *et al.* 2002, 2004). These studies revealed a quite homogeneous pattern of diversity among different northern African Berber populations, with a haplogroup, E-M81 (E1b1b1b in Karafet *et al.* 2008), found at frequencies usually exceeding 65 per cent. The frequency distribution and internal microsatellite diversity of this haplogroup, which is also present in Arabic-speaking populations from northern Africa at lower frequencies (Arredi *et al.* 2004; Bosch *et al.* 2001; Cruciani *et al.* 2002, 2004; Semino *et al.* 2004), have been interpreted as the result of both a demic and a demographic expansion following the neolithization of northern Africa (Arredi *et al.* 2004). Other MSY binary lineages found among the Berbers are present at an average frequency below 5 per cent, and include both “northern African” haplogroups, like E1b1b1a (Cruciani *et al.* 2007) and chromosomes currently found in the Middle East (mainly J1 and G) and sub-Saharan Africa (mainly E1b1a) (Arredi *et al.* 2004; Bosch *et al.* 2001; Cruciani *et al.* 2002).

In the project, we analyzed 29 biallelic MSY markers (Fig. 4.). In all, 16 binary haplogroups have been observed. The “northern African” haplogroup E1b1b1b was the most common haplogroup, with an average frequency of 58.0 per cent. In agreement with previously reported data, this haplogroup was particularly frequent in the Berbers from North-Western Africa (frequencies ranging from 71.0% in the

Middle Atlas Berbers to 86.6% in the Mozabites from Algeria). However, the same haplogroup was detected only in a single individual from the Siwa Oasis (1.1%). The mean frequency of the other 15 haplogroups never exceeded 8 per cent (Table 1). Among these, the “sub-Saharan” haplogroup E1b1a* (frequencies ranging from 3.0% to 7.5%), the “northern African” haplogroup E1b1b1a (1.5%-10.1%) and the “Middle Eastern” haplogroup J1 (1.5%-7.5%) were seen in all the Berber populations analyzed. Conversely, the two commonest haplogroups found in the Siwa Berbers (B2a1a, frequency 28.0%, and R1b1*, frequency 26.9%) were either not observed (B2a1a) or extremely rare (R1b1*) in the Berbers from North-Western Africa (Tab. 1.). Another relatively common haplogroup in Siwa (E1b1b1e, frequency 6.5%) was not detected in the samples of north-western Berbers here analyzed.

The observed pattern of haplogroup frequency distribution resulted in a very high ($F_{ST} = 0.32-0.42$) and significant ($P < 10^{-4}$) degree of genetic differentiation between north-eastern and north-western Berbers. Interestingly, all the haplogroups (B2a1a, R1b1* and E1b1b1e) that differentiate the Egyptian Berbers from the Moroccan and Algerian Berbers have been previously found in sub-Saharan Africa: B2a1a is a quite uncommon haplogroup distributed in a large sub-Saharan area (Cruciani *et al.* 2002); R1b1* is particularly frequent in the central Sahel (Cruciani *et al.* 2002), but has also been found in eastern Africa (Hassan *et al.* 2008), and E1b1b1e was previously observed only in eastern Africa at low frequencies (Cruciani *et al.* 2004). B2a1a and E1b1b1e were never reported in Northern Africa, while R1b1* was observed at low frequencies in Egypt (Luis *et al.* 2004). The Siwa Berbers MSY gene pool turns out to be mainly composed of lineages usually found in sub-Saharan Africa. The Arabic-speaking populations of Northern Africa seem to be characterized by higher frequencies of haplogroups currently found in the Middle East.

Table 1. Frequencies (%) of the Y chromosome haplogroups in five Berber populations

Population (region)	Haplogroup frequency (%)												
	N	A1a	B2a1a	E1a	E1b1a*	E1b1b1a	E1b1b1b	E1b1b1c	E1b1b1e	G	J1	J2	R1a
Asni Berbers (Morocco) ^a	54	-	-	1.9	7.4	1.9	3.7	79.6	-	-	1.9	-	-
Middle Atlas Berbers (Morocco) ^b	69	2.9	-	1.4	4.3	-	10.1	71.0	-	4.3	5.8	-	-
Bouhria Berbers (Morocco) ^a	67	-	-	-	7.5	-	1.5	77.6	-	6.0	1.5	1.5	4.5
Mozabite Berbers (Algeria) ^c	67	-	-	1.5	3.0	-	1.5	86.6	1.5	1.5	1.5	-	-
Siwa Berbers (Egypt) ^a	93	-	28.0	-	6.5	2.2	6.5	1.1	2.2	3.2	7.5	6.5	-
Total	350	0.6	7.4	0.9	5.7	0.9	4.9	58.0	0.9	1.7	3.1	2.0	0.9

^a These samples were previously analyzed for the M78 marker (Cruciani *et al.* 2007).

^b A subset of this sample was previously analyzed for some of the markers here studied (Cruciani *et al.* 2002).

^c A subset of this sample was previously analyzed for the M78 marker (Cruciani *et al.* 2007).

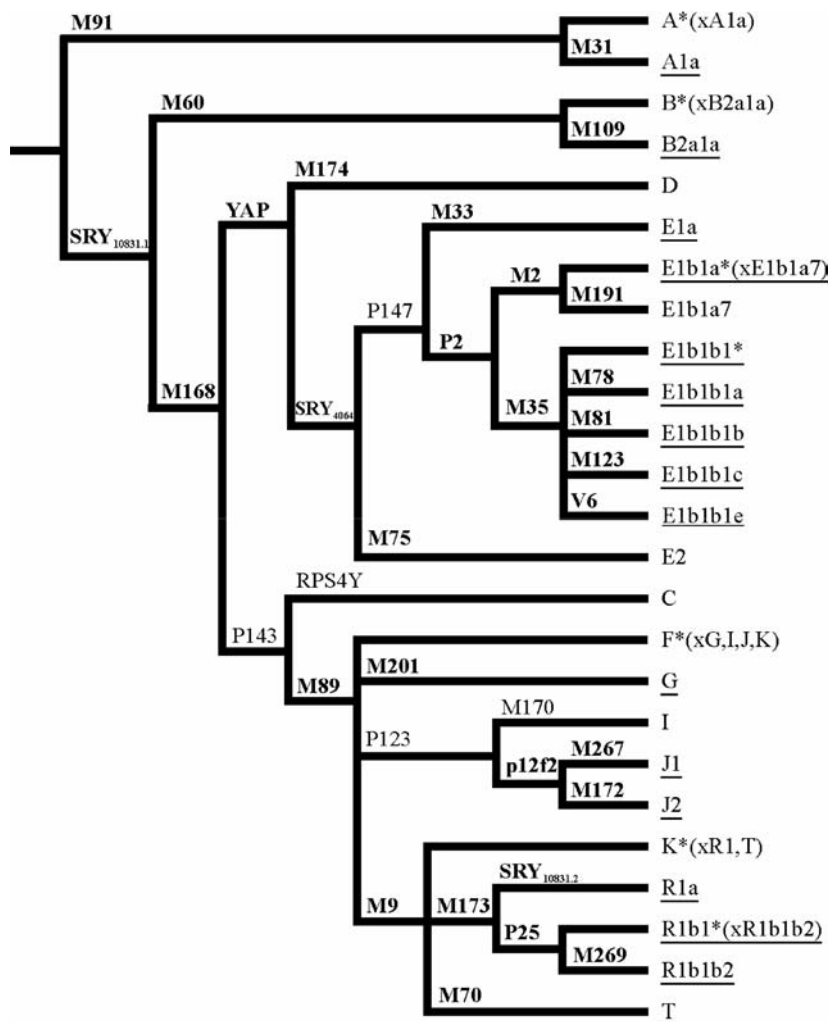


Figure 4. A simplified version of the Y chromosome haplogroup tree. Markers analyzed in this study are indicated in bold. Observed haplogroups are underlined. Haplogroup nomenclature is as reported in Karafet *et al.* (2008)

Synthesis

Biological studies were conducted on various genetic markers and on many samples. Results have been used to (i) describe the genetic diversity of Berber populations, (ii) assess the genetic affinities among North Africans (Berbers vs. Berbers and

Berbers vs. Arabs), and (iii) connect the current genetic diversity of Berber-speakers with the several past migrations which took place in North Africa.

As previously published, and for the four polymorphic systems here studied (GM allotypes, Alu insertions, mtDNA, and MSY), we confirm that North-African Berber genetic pools are highly diversified but with the same overall structure. Indeed, with the exception of the Y chromosome polymorphisms, they are characterized by high frequencies of markers commonly observed in Europeans and by a lower frequency of Sub-Saharan markers. Moreover, they show relatively high frequencies of polymorphisms that seem to be specific to Northern Africa (the U6 mitochondrial haplogroup, the CD4 110(-) and DM 107(-) Alu haplotypes, and the E1b1b1b Y haplogroup).

How can we explain these results? On the one hand, the genetic proximity observed between the Berbers and southern Europeans reveals that these groups shared a common ancestor. Two hypotheses are discussed: one would date these common origins to the Upper Palaeolithic with the expansion of anatomically modern humans, from the Near East to both shores of the Mediterranean Sea; the other supports the Near-Eastern origin, but would rather date it to the Neolithic, around 10,000 years ago (Ammerman and Cavalli-Sforza 1973; Barbujani *et al.* 1994; Myles *et al.* 2005; Rando *et al.* 1998). Polymorphisms common to Berbers and south-Europeans could also have been introduced or supported by genetic flows through the Straits of Gibraltar. For example, genetic exchanges could have taken place during prehistory, while European populations expanded from the Franco-Cantabrian refuge, 15,000 years ago (Achilli *et al.* 2005; Cherni *et al.* 2008; Ennafaa *et al.* 2009; Housley *et al.* 1997; Pereira *et al.* 2005). Or these gene flows could have occurred during history, with the occupation during nearly seven centuries (from the 8th to the 15th century) of the Iberian Peninsula by Almoravide, and then Almohade Muslim Berber troops. On the other hand, the differentiation observed between North-Africans and sub-Saharan populations shows that settlement of these areas happened in different migration waves. The Berber genetic heritage also consists of relatively high frequencies of haplogroups originating from various parts of sub-Saharan Africa. For example for mitochondrial haplogroups: the L0a, L3i, and L4 clades are from East-Africa, L1b, L2b, and L3b are from West Africa, and L3e originated in the Sudan. Another example for MSY: E1b1a and B2a1a originated in Central/Western Africa, while E1b1b1e is common in Eastern Africa. Thus, it poses a question about the Sahara desert's role in population movements and exchanges. The Saharan areas separate the African continent into two distinct regions and this geographical barrier has doubtlessly influenced the peopling of Northern Africa by compelling human groups to reach the Mediterranean coast through an East-West route. It should be specified that the Sahara was not always a desert, and that it underwent enormous variation between wet and dry

periods, offering green spaces favourable for human occupation and animal domestication (Aumassip *et al.* 1994; Said and Faure 1990). Thus, there were probably exchanges between African prehistoric populations, exchanges during which markers predominant in sub-Saharan groups would have been introduced into the Berber gene pool. Contacts between Northern Africa and the great sub-Saharan empires (such those of Ghana, of Mali, or the Songhai empire) have certainly played a role, and, even more so, the long-lasting Trans-Saharan trade of gold, salt and slaves. Therefore, it is clear that neither the Strait of Gibraltar nor the Sahara desert appears to have halted the movements and exchanges between Southern European, Northern African, and sub-Saharan populations. Since prehistory, this interaction has taken place in a human and genetic framework which was already diversified, and the ancestors of the Berbers were able to exchange some markers with surrounding populations with a different culture and genetic pool.

Even though in an overall analysis Berber populations may be considered to constitute a relatively homogeneous group, the case is very different when one focusses on the internal diversification of the North-African Berber groups. Indeed, by collecting samples from the extreme East of the Berber-speakers distribution area, we could highlight a clear and significant differentiation between Berbers from the Maghreb and Egyptian Berbers. While the first are related to European populations (because of, e.g., high frequency of European GM haplotypes and H mitochondrial haplogroups), a consistent portion of the gene pool of Siwa Berbers was shared with East- and Sub-Saharan populations (i.e. M1, L0a, L3i, L4*, and L4b mitochondrial haplogroups and E1b1b1e, B2a1a, R1b1* Y lineages), and with Near-Easterners (J1 Y haplogroup).

The differentiation observed between North-Western and North-Eastern African Berbers seems to be correlated with linguistic studies which show that Siwan has a number of features which are only encountered in this oasis, and not elsewhere in Berber. In fact, Siwan poses intricate problems from the point of view of a historical linguist. On the one hand, in many respects Siwan is quite similar to the Berber languages in central Libya (El Foqaha, Sokna, Djebel Nefusa) and even further west (cf. Kossmann 1999: 31). At this point it differs from its immediate neighbour to the west, Awdjilah Berber, which diverges from all other Berber languages of the region, and which may constitute an early branching of the Berber language family (cf. Naumann 1999). On the other hand, Siwa has a number of features which are either unique in Berber, or which are only encountered in Siwa and in Awdjilah. Some of these features are the loss of the subject-relative form (so-called participle); the loss of the deictic clitics to the verb; the fixed position of verbal clitics after the verb, in contrast to other Berber languages, where the position of the clitics with respect to the verb varies according to syntactic conditions; the introduction of the 3sg:m prefix *y-* ~ *i-* in the 3pl of the conjugated verb,

i.e. *i-fetk-en* 'they opened', where other Berber languages would rather have *fetk-en*. (An alternative explanation would be that the 3pl form is derived from the ancient participle, which, in other Berber languages, has the same form.) The interesting point about these unique features is that most of them seem to constitute systemic simplifications. This suggests an explanation in the sense of the sociolinguistic history of Siwa (and possibly Awdjilah). While language change is paramount in any speech community, and simplifications may occur under all social circumstances, large-scale simplifications are often supposed to be due to a large influx into the speech community of imperfect speakers. When this imperfect speech variety gains momentum, either because of the social position of the non-native speakers, or simply because of their numbers, some of these simplifications may become part of the language as a whole (cf. Kusters 2003). One has the impression that this is what happened in Siwa (and maybe also in Awdjilah): apparently elements of non-native speech have become normative in the language.

The differentiation would have been the result of a conjunction of several geographical, prehistoric, and historic factors which guided contacts (and thus exchanges) between local and migrating populations. First, in addition to the geographical distance, that certainly increases the genetic distance, the geographical location of Berber populations is very peculiar: the Berbers from the Maghreb are at the end of a long migration route, whereas Berbers from Siwa are rather in a crossroads between the Middle East, East Africa, sub-Saharan areas and the North-African corridor. Meetings and exchanges between local and migrating populations were certainly not identical in North-Western and North-Eastern Africa. Then, the prehistory and history of the populations from the Maghreb are different from those of the Egyptian group. For the Siwa oasis, there is very little information on the exact prehistoric period which was the starting point of Berber culture. However we know that throughout history the oasis was called at by successive human groups, like pilgrims travelling to Mecca, Mediterranean traders, or Sahelian slave merchants. Siwa may even have been repopulated at a certain moment in time by Libyan Berber-speakers driven from their land by Arab conquerors. Lastly, it experienced a period of decline and faced between the ninth and twelfth century AD a drastic demographic decline (Fakhry 1973). Therefore, the current gene pool of Siwa people could be related to recent migrations/founder effects or it could be the result of the various genetic exchanges which occurred in the Past.

Although Northern African history is quite well-known, it only provides very little (if any) evidence for demographic aspects. However, some data exist concerning the Arab conquest which had the most important impact on North African demography: the first migration (7th century) implied only a few thousands of individuals, but the second invasion (11th century) was massive and many Bedouins invaded North-Western Africa (McEvedy 1980). Therefore it is legitimate

to think that, in addition to their cultural impact, that Arab populations left a footprint in the Berber genetic pool. Surprisingly (or not?), our analyses of GM allotypes, Alu sequences, mitochondrial and Y haplogroups did not show a strong genetic differentiation between Berber-speakers and Arabic-speaking population. Under these circumstances, it is thus clear that a purely linguistic or cultural definition of Berbers cannot be correlated with a genetic unit of population. How shall we interpret this? As North African Berber-speakers and Arabic-speakers are not genetically differentiated, it may suggest that they share a common Berber ancestry. Thus, the differences observed today between Berbers and Arabs would be purely linguistic and might be explained by the Berber ancestors reacting in different ways to the Arab conquest: for some groups it may have been a cultural event that changed their language and religion, but had a weak genetic impact on later generations; they would be the present-day Arabs. On the other hand, the predecessors of current Berbers resisted Arabization, and their descendants, living in more isolated regions, maintain their linguistic identity to this day. Such a hypothesis supposes that the current Arabic- and Berber-speaking populations had similar gene pools before the Arab conquest. Nevertheless, this does not necessarily solve the question altogether. In fact, only the analysis of other Arab populations from North Africa and from the Arab Peninsula could answer this. Certainly, the Arab conquest was mainly a cultural phenomenon, leading to religious and linguistic changes, but it must also have left a genetic print in the North-African populations. Finally, since we cannot assess the genetic contribution of Arab invaders to North-African Berbers, it is reasonable to assume that the current Berber gene pool is the result of reciprocal genetic exchanges between past populations. Through the analysis we conducted, we conclude that the origins and diversity of Berber populations are old and complex.

Acknowledgements

We are grateful to all the people and infrastructures who contributed to the collecting of blood samples: F. El Chennawi, A. Baali, M. Cherkaoui and M. Melhaoui for their help in the collection of the Moroccan and Egyptian samples. We also particularly thank for their help in the genetic analyses: Evelynne Guitard, André Sevin, and the team of the Genopole Toulouse platform from Auzerville (<http://genopole-toulouse.prd.fr/>) for their technical help. Acknowledgements to G. Philippson (Laboratoire Dynamique du Langage, CNRS UMR 5596, Lyon, France) who contributed with N.L. to the linguistic approach of the new Berber populations studied in this project. This work was made within the framework of the E.C. Sixth Framework Programme under Contract no. ERAS-CT-2003-980409

(EUROCORES project of the European Science Foundation) (to J.M.D.). It benefited from funding from Progetti Ricerca Interesse Nazionale 2007 (Italian Ministry of the University) (to A.T.), Sapienza Università di Roma (Grandi Progetti Ateneo 2007) and the Italian Ministry of the University (Progetti di Ricerca di Interesse Nazionale n. 20073RH73W_002) (to R.S.), CNRS (Centre National de la Recherche Scientifique) (to J.M.D.), Université Paul Sabatier Toulouse III, and the Conseil Régional de Midi-Pyrénées (Toulouse, France) (to J.M.D.).

References

- Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., Cruciani, F., Zeviani, M., Briem, E., Carelli, V., Moral, P., Dugoujon, J.M., Roostalu, U., Loogväli, E.L., Kivisild, T., Bandelt, H.J., Richards, M., Villems, R., Santachiara-Benerecetti, A.S., Semino, O., Torrioni, A. 2004. "The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool". *American Journal of Human Genetics* 75(5): 910–918.
- Achilli, A., Rengo, C., Battaglia, V., Pala, M., Olivieri, A., Fornarino, S., Magri, C., Scozzari, R., Babudri, N., Santachiara-Benerecetti, A.S., Bandelt, H.J., Semino, O., Torrioni, A., 2005. "Saami and Berbers—an unexpected mitochondrial DNA link". *American Journal of Human Genetics* 76(5): 883–886.
- Ammerman, A. and Cavalli-Sforza, L.L., 1973. "A population model for the diffusion of early farming in Europe", in *The explanation of culture change: models in prehistory*, C. Renfrew (ed), 348–358. Duckworth, London.
- Arredi, B., Poloni, E.S., Paracchini, S., Zerjal, T., Fathallah, D.M., Makrelouf, M., Pascali, V.L., Novelletto, A., Tyler-Smith, C., 2004. "A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa". *American Journal of Human Genetics* 75: 338–345.
- Athanasiadis, G., Esteban, E., Via, M., Dugoujon, J.M., Moschonas, N., Chaabani, H., Moral, P., 2007. "The X chromosome Alu insertions as a tool for human population genetics: data from European and African human groups". *European Journal of Human Genetics* 15(5): 578–583.
- Aumassip, G., Ferhat, N., Heddouche, A., Vernet, R., 1994. "Le milieu saharien aux temps préhistoriques", in *Milieus, hommes et techniques du Sahara préhistorique. Problèmes actuels*, G. Aumassip, N. Ferhat, A. Heddouche (eds), 9–30. Paris, L'Harmattan.
- Avice, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reece, C.A., Saunders, N.C., 1987. "Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics". *Annual Review of Ecology and Systematics* 18: 489–522.
- Barbujani, G., Pilastro, A., De Domenico, S., Renfrew, C., 1994. "Genetic variation in North Africa and Eurasia: Neolithic demic diffusion vs. Palaeolithic colonisation". *American Journal of Physical Anthropology* 95: 137–154.
- Blench, R., 2001. "Types of Language Spread and their Archaeological Correlates: The Example of Berber". *Origini* 23: 169–190.
- Bosch, E., Calafell, F., Comas, D., Oefner, P.J., Underhill, P.A., Bertranpetit, J., 2001. "High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula". *American Journal of Human Genetics* 68: 1019–1029.

- Brakez, Z., Bosch, E., Izaabel, H., Akhayat, O., Comas, D., Bertranpetit, J., Calafell, F., 2001. "Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area". *Annals of Human Biology* 28: 295–307.
- Brett, M. and Fentress, E., 1996. *The Berbers*. Oxford: Blackwell Publishers.
- Cavalli-Sforza, L.L., Menozzi, P., Piazza, A., 1994. *The history and geography of human genes*. Princeton: Princeton University Press.
- Chaabani, H., Helal, A.N., Van Loghem, E., Langaney, A., Benammar Elgaaied, A., Rivat Peran, L., Lefranc, G., 1984. "Genetic study of Tunisian Berbers. I. Gm, Am and Km immunoglobulin allotypes and ABO blood groups". *Journal of Immunogenetics* 11: 107–113.
- Chaabani, H. and Cox, D.W., 1988. "Genetic characterization and origin of Tunisian Berbers". *Human Heredity* 38: 308–316.
- Chaker, S., 1995. *Linguistique berbère: études de syntaxe et de diachronie*. Paris/ Louvain: Peeters.
- Cherni, L., Loueslati, B.Y., Pereira, L., Ennafaa, H., Amorim, A., El Gaaied, A.B., 2005. "Female gene pools of Berber and Arab neighboring communities in central Tunisia: microstructure of mtDNA variation in North Africa". *Human Biology* 77(1): 61–70.
- Cherni, L., Fernandes, V., Pereira, L., Costa, M.D., Goios, A., Frigi, S., Yacoubi-Loueslati, B., Amor, M.B., Slama, A., Amorim, A., El Gaaied, A.B., Pereira, L. 2008. "Post-last glacial maximum expansion from Iberia to North Africa revealed by fine characterization of mtDNA H haplogroup in Tunisia." *American Journal of Physical Anthropology* 139(2): 253–260
- Côrte-Real, H.B., Macaulay, V.A., Richards, M.B., Hariti, G., Issad, M.S., Cambon-Thomsen, A., Papiha, S., Bertranpetit, J., Sykes, B.C., 1996. "Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis". *Annals of Human Genetics* 60(4): 331–350.
- Coudray, C., 2006. *Histoire génétique et évolution des populations berbérophones nord-africaines*. PhD thesis, Toulouse III University, France.
- Coudray, C., Guitard, E., Kandil, M., Harich, N., Melhaoui, M., Baali, A., Sevin, A., Moral, P., Dugoujon, J.M., 2006. "Study of GM immunoglobulin allotypic system in Berbers and Arabs from Morocco". *American Journal of Human Biology* 18: 23–34.
- Coudray, C., Olivieri, A., Achilli, A., Maria Pala M., Melhaoui M., Cherkaoui, M., El-Chennawi, F., Kossmann, M., Torroni, A., Dugoujon, J.M. 2009. "The complex and diversified mitochondrial gene pool of Berber populations". *Annals of Human Genetics* 73(2):196–214.
- Coudray, C., Guitard, E., El Chennawi, F., Dugoujon, J.M. 2008. "Study of GM immunoglobulin allotypes in Egyptian Berbers (Siwa Oasis)". In *Egypt at its Origins 2*, Midant-Reynes B. and Tristant Y. (eds) 172: 265–273. Orientalia Lovaniensa Analecta.
- Cruciani, F., Santolamazza, P., Shen, P., Macaulay, V., Moral, P., Olckers, A., Modiano, D., Holmes, S., Destro-Bisol, G., Coia, V., Wallace, D.C., Oefner, P.J., Torroni, A., Cavalli-Sforza, L.L., Scozzari, R., Underhill, P.A., 2002. "A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes". *American Journal of Human Genetics* 70:1197–1214.
- Cruciani, F., La Fratta, R., Santolamazza, P., Sellitto, D., Pascone, R., Moral, P., Watson, E., Guida, V., Beraud Colomb, E., Zaharova, B., Lavinha, J., Vona, G., Aman, R., Cali, F., Akar, N., Richards, M., Torroni, A., Novelletto, A., Scozzari, R., 2004. "Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa". *American Journal of Human Genetics* 74:1014–1022.
- Cruciani, F., La Fratta, R., Trombetta, B., Santolamazza, P., Sellitto, D., Beraud Colomb, E., Dugoujon, J.M., Crivellaro, F., Benincasa, T., Pascone, R., Moral, P., Watson, E., Melegh, B., Barbujani, G., Fuselli, S., Vona, G., Zagradsnik, B., Assum, G., Brdicka, R., Kozlov, A.I.,

- Efremov, G.D., Coppa, A., Novelletto, A., Scozzari, R., 2007. "Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12". *Molecular Biology and Evolution* 24: 1300–1311.
- Dugoujon, J.M., Hazout, S., Loirat, F., Mourrieras, B., Crouau-Roy, B., Sanchez-Mazas, A., 2004. "GM haplotype diversity of 82 populations over the world suggests a centrifugal model of human migrations". *American Journal of Physical Anthropology* 125: 175–192.
- Dugoujon, J.M., Lemaire, O., Guitard, E., Sevin, A., Larrouy, G., Moral, P., Sabir, B., Baali, A., Cherkaoui, M., 2005. "Diversité des allotypes des immunoglobulines (systèmes Gm et Km) d'une population berbère de la vallée du Tacchedirt (Haut Atlas, Maroc). Comparaison à d'autres populations africaines et européennes et histoire du peuplement de l'Afrique du Nord". *Bulletin et Mémoires de la Société d'Anthropologie de Paris* 17:57–72.
- Ennafaa, H., Cabrera V.M., Abu-Amero K.K., Gonzalez A.M., Amor, M.B., Bouhaha, R., Dzimiri, N., Elgaaied, A.B., Larruga, J.M., 2009. "Mitochondrial DNA haplogroup H structure in North Africa". *B.M.C. Genetics* Feb 25; 10:8 (10 pages)
- Esteban, E., González-Pérez, E., Harich, N., López-Alomar, A., Via, M., Luna, F., Moral, P., 2004. "Genetic relationships among Berbers and South Spaniards based on CD4 microsatellite/Alu haplotypes". *Annals of Human Biology* 31(2): 202–212.
- Esteban, E., Via, M., González-Pérez, E., Santamaría, J., Dugoujon, J.M., Vona, G., Harich, N., Luna, F., Saetta, A.A., Bissar, N., Moral P., 2005. "An unexpected wide population variation of the G1733A polymorphism of the androgen receptor gene: data on the Mediterranean region". *American Journal of Human Biology* 17(6): 690–695.
- Esteban, E., Rodon, N., Via, M., Gonzalez-Perez, E., Santamaria, J., Dugoujon, J.M., Chennawi, F., Melhaoui, M., Cherkaoui, M., Vona, G., Harich, N., Moral, P., 2006. "Androgen receptor CAG and GGC polymorphisms in Mediterraneans: repeat dynamics and population relationships". *Journal of Human Genetics* 51(2): 129–136.
- Fadhlaoui-Zid, K., Dugoujon, J.M., Elgaaied, A.B., Ben Amor, M., Loveslati, B.Y., Khodjet El Khil, H., Sanchez-Mazas, A., 2004a. "Genetic diversity in Tunisia: a study based on the Gm polymorphism of human immunoglobulins". *Human Biology* 76: 559–567.
- Fadhlaoui-Zid, K., Plaza, S., Calafell, F., Ben Amor, M., Comas, D., Bennamar El gaaied, A., 2004b. "Mitochondrial DNA heterogeneity in Tunisian Berbers". *Annals of Human Genetics* 68(3), 222–233.
- Fakhry, A., 1973. *Siwa Oasis*. The American University in Cairo Press.
- Gonzalez-Perez, E., Moral, P., Via, M., Vona, G., Varesi, L., Santamaria, J., Gaya-Vidal, M., Esteban, E., 2007. "The ins and outs of population relationships in west-Mediterranean islands: data from autosomal Alu polymorphisms and Alu/STR compound systems". *Journal of Human Genetics* 52(12), 999–1010.
- Hachid, M., 2000. *Les premiers Berbères: entre Méditerranée, Tassili et Nil*. Aix-en-Provence: Edisud & Alger: Ina-yas.
- Hassan, H.Y., Underhill, P.A., Cavalli-Sforza, L.L., Ibrahim, M.E., 2008. "Y-chromosome variation among Sudanese: Restricted gene flow, concordance with language, geography, and history". *American Journal of Physical Anthropology* 137(3):316–323
- Housley, R.A., Gamble, C.S., Street, M., Pettitt, P., 1997. "Radiocarbon evidence for the late glacial human recolonisation of northern Europe". *Proceedings of the Prehistoric Society* 63: 25–54.
- Karafet, T.M., Mendez, F.L., Meilerman, M.B., Underhill, P.A., Zegura, S.L., Hammer, M.F., 2008. "New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree". *Genome Research* 18: 830–838.

- Kéfi, R., Stevanovitch, A., Bouzaid, E., Béraud-Colomb, E., 2005. "Diversité mitochondriale de la population de Taforalt (12.000 ans BP – Maroc): une approche génétique à l'étude du peuplement de l'Afrique du Nord". *Anthropologie* 43: 1–11.
- Kossmann, M.G., 1997. *Grammaire du parler berbère de Figuig*. Louvain & Paris: Peeters.
- Kossmann, M.G., 1999. *Essai sur la phonologie historique du berbère*. Köln: Rüdiger Köppe.
- Krings, M., Salem, A.E., Bauer, K., Geisert, H., Malek, A.K., Chaix, L., Simon, C., Welsby, D., Di Rienzo, A., Utermann, G., Sajantila, A., Paabo, S., Stoneking, M., 1999. "mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration?" *American Journal of Human Genetics* 64(4): 1166–1176.
- Kusters, W. 2003. *Linguistic Complexity. The Influence of Social Change on Verbal Inflection*. Utrecht: LOT.
- Laoust, E., 1931. *Siwa. I. Son parler*. Paris: Leroux.
- Lefèvre-Witier, P., 1982. *Structure et dynamique génétiques d'une communauté rurale du Sahara central (Ideles)*. PhD thesis, Toulouse III University, France.
- Lefranc, G., De Lange, G., Rivat, L., Langaney, A., Lefranc, M.P., Ellouze, F., Sfar, G., Sfar, M., Van Loghem, E., 1979. "Gm, Am and Km immunoglobulin allotypes of two populations in Tunisia". *Human Genetics* 50: 199–211.
- Lefranc, M.P., Lefranc, G., 1990. "Molecular genetics of immunoglobulin allotype expression", in *The Human IgG subclasses: molecular analysis of structure, function and regulation*, Shakib (ed), 43–78. Oxford: Pergamon Press.
- Louali, N., Philippson, G., 2004. "Berber expansion into and within north-west Africa: a linguistic contribution". *Afrika und Übersee* 87: 105–129.
- Louali, N., Philippson, G., 2005. "Deux systèmes accentuels berbères: le siwi et le touareg. *Faits de langue*" 26(1): 11–22.
- Loveslati, B.Y., Sanchez-Mazas, A., Ennafaa, H., Marrakchi, R., Dugoujon, J.M., Lefranc, J.M., Elgaaied, A.B., 2001. "A study of Gm allotypes and immunoglobulin heavy gamma IGHG genes in Berbers, Arabs and sub-Saharan Africans from Jerba Island, Tunisia". *European Journal of Immunogenetics* 28: 531–538.
- Loveslati, B.Y., Cherni, L., Khodjet-Elkhil, H., Ennafaa, H., Pereira, L., Amorim, A., Ben Ayed, F., Ben Ammar Elgaaied, A., 2006. "Islands inside an island: reproductive isolates on Jerba island". *American Journal of Human Biology* 18(1): 149–153.
- Luis, J.R., Rowold, D.J., Regueiro, M., Caeiro, B., Cinnioglu, C., Roseman, C., Underhill, P.A., Cavalli-Sforza, L.L., Herrera, R.J., 2004. "The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations". *American Journal of Human Genetics* 74: 532–544.
- Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B., Torroni, A., 1999. "The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs". *American Journal of Human Genetics* 64(1): 232–249.
- McEvedy, C., 1980. *The Penguin Atlas of African history*. New York: Penguin books.
- Myles, S., Bouzekri, N., Haverfield, E., Cherkaoui, M., Dugoujon, J.M., Ward, R., 2005. "Genetic evidence in support of a shared European-North African dairying origin". *Human Genetics* 117: 34–42.
- Naumann, C., 1999. *Vergleich demonstrativer Formative ausgewählter Berbersprachen*. Leipzig: Institut für Afrikanistik.
- Olivieri, A., Achilli, A., Pala, M., Battaglia, V., Fornarino, S., Al-Zahery, N., Scozzari, R., Cruciani, F., Behar, D.M., Dugoujon, J.M., Coudray, C., Santachiara-Benerecetti, A.S., Semino,

- O., Bandelt, H.J., Torroni, A., 2006. "The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa". *Science* 314(5806): 1767–1770.
- Pereira, L., Richards, M., Goios, A., Alonso, A., Albarrán, C., Garcia, O., Behar, D.M., Gölge, M., Hatina, J., Al-Gazali, L., Bradley, D.G., Macaulay, V., Amorim, A., 2005. "High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium". *Genome Research* 15(1): 19–24.
- Plaza, S., Calafell, F., Helal, A., Bouzerna, N., Lefranc, G., Bertranpetit, J., Comas, D., 2003. "Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean". *Annals of Human Genetics* 67: 312–328.
- Rando, J.C., Pinto, F., Gonzalez, A.M., Hernandez, M., Larruga, J.M., Cabrera, V.M., Bandelt, H.J., 1998. "Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European near-eastern and sub-Saharan populations". *Annals of Human Genetics* 62(6): 531–550.
- Said, R., Faure, H., 1990. "Chronological framework: African pluvial and glacial epochs". In *General history of Africa*, J. Ki-Zerbo (ed), 146–166. Paris: Unesco.
- Semino, O., Magri, C., Benuzzi, G., Lin, A.A., Al-Zahery, N., Battaglia, V., Maccioni, L., Triantaphyllidis, C., Shen, P., Oefner, J.P., Zhivotovsky, L.A., Roy, K., Torroni, A., Cavalli-Sforza, L.L., Underhill, P.A., Santachiara-Benerecetti, A.S., 2004. "Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area". *American Journal of Human Genetics* 74: 1023–1034.
- Steinberg, A.G., Cook, C.E., 1981. *The distribution of Human Immunoglobulin Allotypes*. Oxford: Oxford University Press.
- Stevanovitch, A., Gilles, A., Bouzaid, E., Kefi, R., Paris, F., Gayraud, R.P., Spadoni, J.L., El-Chenawi, F., Beraud-Colomb, E., 2004. "Mitochondrial DNA sequence diversity in a sedentary population from Egypt". *Annals of Human Genetics* 68(1): 23–39.
- Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Mirazón Lahr, M., Foley, R.A., Oefner, P.J., Cavalli-Sforza, L.L., 2001. "The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations". *Annals of Human Genetics* 65: 43–62.
- Vycichl, W., 2005. *Berberstudien & A Sketch of Siwi Berber (Egypt)*, D. Ibrisimow and M.G. Kossmann (eds), Köln: Rüdiger Köppe.

East meets West

The Assam corridor

François Jacquesson

The paper describes the ethnic and linguistic situation in North-East India, a corridor between highland Tibet and lowland Bangladesh, with currently more than 100 languages. The paper pictures the mosaic of languages and the logic of contacts in the region.

Introduction: the region and the problem

We had decided to explore what the corridor just south of the Himalayan range had to offer to from our perspective. The situation is different from the wide and open steppes to the north. The narrow Assam corridor in the north-easternmost corner of South Asia is specific in many ways. Yet it offered an opportunity for the exploration of how human groups interact and organize themselves in this kind of bottleneck.

The lowest part (Figure 1) of the region is Bengal. It is formed from the alluvia of two hydrographic basins: the Ganges' on the west, and the Brahmaputra's on the east. The latter flows down from the Tibetan plateau, crosses the easternmost Himalaya and in its middle course, while running south-west, forms the Assam plain with its first deposits. After a bend around the Meghalaya hills, the river runs southwards and joins the Ganges.

The Assam plain looks like a finger pointing north-east into the high mountains. Many ethnic groups or populations are found in this rather small area where about fifty languages are currently being spoken, and twice as much if one includes the surrounding mountain slopes.

The aim of the project was to enquire into how these groups were formed and how they evolved, if their diversity was inherited or if it was locally developed, and finally with which criteria to evaluate this diversity.

A basic approach would be altitude, with its ecological implications. It seems natural that the open space in the plains, with water available for growing rice (the

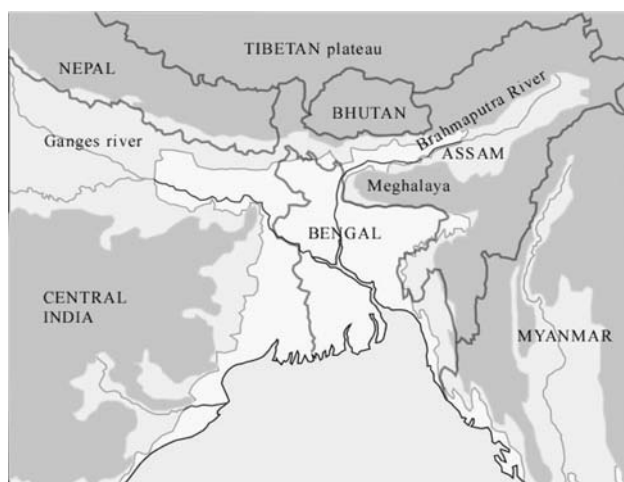


Figure 1. East and North-east of the Indian subcontinent.
With altitude (colours) and political borders.

staple food of the whole continent), can sustain a denser population than the difficult mountain slopes around; and this is what we find. However, it turns out (a) that population density differs significantly within the too broad category of «mountains», (b) that political systems are a significant parameter to explain group size, (c) that linguistic mapping does not completely explain the group fragmentation, nor conversely.

A/ Population density, its causes and side-effects

Although the demography of the Assam plains and of the surrounding hills has been much transformed, a knowledge of the history of Assam, of the successive steps of its colonization, and of the ways in which statistical sources were compiled, allow us to advance in our enquiries.

The best tool is the 1931 Census, the more precise and the last of the British censuses. Because political borders have changed and ethnic names are to be analysed with care, the whole treatment is delicate and lengthy, and we will here only sketch the method and give the results.

The Census data (Figure 2) have to be reformulated in compatible measurements, and separated from their political context, which implies for instance knowing that Khasi Sates (nominally independent at that time) are comparable with «Khasi & Jaintia» (which were not) etc. «Surma Valley» is a category for the six groups that follow, so is «Assam Valley» for the next seven groups. «Sadiya Frontier Tract» and «Balipara Frontier Tract» are border areas under British Control

	km ²	villages	persons (rural)	pers. /km ²	km ² / village	pers. /vill
				a	b	c
					6,2	
					1,2	228,7
					6,6	104,9
					23,1	370,2
					38,8	230,4
					3,2	270,2
					2,9	342,5
					3,7	289,6
					4,3	236,4
					5,9	398,5
					4,4	278,9
					3,8	88,4
					22,2	131,3
					38,2	135,5
					16,3	262,2
					6,5	110,7

Figure 2. An example of data processing of the 1931 Census of Assam. Colours show the density layers

(a typical feature of the late «Great Game») and lie inside what is now Arunachal Pradesh. Manipur State is a nominal rajahship, and a passage to Burma/Myanmar.

The analysis shows diagnostic population density layers. The lowest layer is in pale blue (< 7 hb/km²), the highest layer is violet (> 55 hb/km²), with an interesting medium layer, distinct from the other two (between 16 and 24 hb/km²).

This is summed up in Figure 3 (below).

We are not surprised to see that the higher densities, here in red, correspond to the lowland populations (Nowgong, Cachar, Lakhimpur, Sibsagar, Darrang and Goalpara districts), while the lower densities, in blue, correspond to the hills inhabited by «Tribals» (Balipara, Lushai, Sadiya, Naga, Khasi, Khasi States, Garo). Two observations appear fundamental at this stage.

1. There are two distinct levels of low densities. One is very low and «blue» in Figure 2; the Tani case is emphasized. The other one is medium low and «orange» in Figure 2; the Naga case is emphasized.
2. The two main categories, in blue and red in Figure 3, are labelled «Tribal» and «Non-Tribal» respectively, because this would be the assessment of the census officers or, for that matter, of many local persons perhaps still today. An assessment of the same sources would be that People in the plains are not Tribal (it is administratively not true, as we shall see), while Mountain People are Tribal. These two points are examined in the next section.

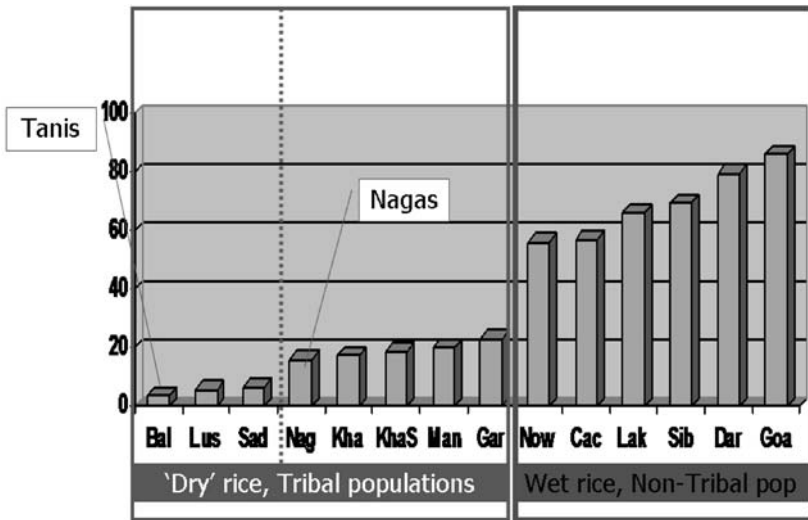


Figure 3. Population density typology in Assam, 1931 Census

B/ Politics, identity and ecology

The first remark on the contrast between low and very low densities is all important when we analyse pre-industrial communities and economics. In order to describe the contrast more vividly, we will use two contrasting types: the Tani population in Arunachal Pradesh (of which Balipara Frontier Tract in Figs. 2 and 3 is an example) and the Naga population in Nagaland. Our treatment here is simplified, but the main features of the contrast between the two groups are not.

The two groups (Figure 4) live on either side of the Brahmaputra river. The Tani group lives on the northern side, in the easternmost Himalayas where one can also find a Mishing group; the Naga groups live on the southern side, astride the range between India and Myanmar; the case of the Karbi-Arleng tribe which lives in the Mikir Hills in Assam (the small hills at the foot of the M of ASSAM on Figure 4) is not discussed here.

The contrast is easily described. Tani groups were sparse, linked to one another in a loose but efficient way, and population density was very low; villages were very small and tended to move. Naga groups were very different from each other in customs and language, brutally hostile to each other, and dense for a mountain population (Figure 5); villages were impressively big and built as fortresses. Landscapes are not identical in the Tani and Naga regions and geographers would find many differences; it is also true that fields are more common and more concentrated in the Naga country. However, both regions are of about the same size,

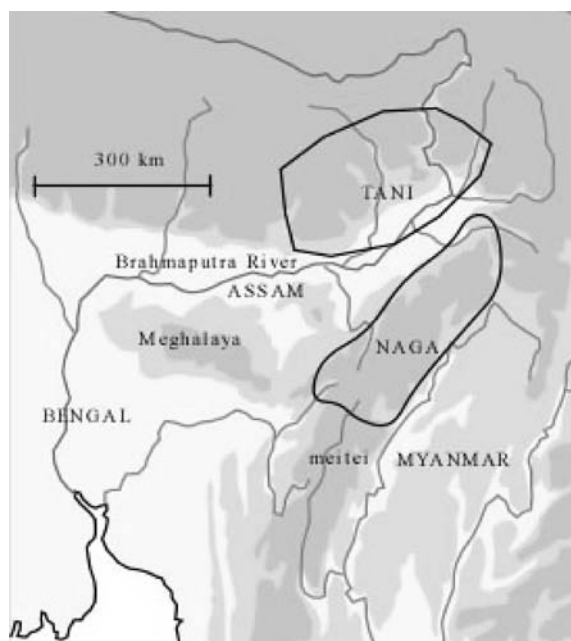


Figure 4. Location of the Tani and Naga groups.



Figure 5. An Ao Naga village, 1946

and practise swidden cultivation, although much varied in patterns among the Nagas. Therefore, it seems reasonable not to overlook social organisation and political systems which play a role in setting rules for common life, in using common resources and in providing, so to speak, reasons for feuds.

Languages provide an interesting parameter. Most linguists who studied these two regions agree on stressing the contrast between the dialect-continuum pattern of the Tani regions, and the deliberate production of self-contained and parochial languages among the Naga groups.

A dialect continuum is an embarrassing fact for politicians or census people. It means that a given village speech is perfectly understood in the next village, still well understood farther on, rather poorly understood when getting farther, and completely incomprehensible when you keep on travelling from it. There is no border within the area, because between neighbour villages understanding is easy, although it is not easy at all when people from two external sides happen to meet. It is a smooth gradation. With time and perhaps the weakening of older family ties, a dialect border may form, but as long as the continuum is maintained, no border can be pointed out.

The situation among the Nagas is strikingly different, because a century and a half ago, if you crossed the border, you might be killed; consequently, most people knew where the border was. Head hunting was rampant – the standard adjective in British colonial ethnography. The big villages on their hill crests, with their massive stone walls and their murderous hidden snares on the paths, made possible the development of very different modes of speaking: languages diverged fast.

Therefore, the number of different languages and their concentration are not, in such a case, the result of a longer period of «natural development and diversification». Whatever the previous situation (provided that «previous» here has more meaning than «primitive»), it is clear that the language map of the Naga country, a great number of distinct patches, has more to do with the deliberate production or stressing of distinctions. The high reputation of the «Noble Naga» among British officers and ethnographers shows that the sociologist Pierre Bourdieu was not wrong when he described in *La Distinction* how the aptitude for nobility was a matter of producing distinct features non-stop (Bourdieu 1979).

In current ideology, the feature «Tribal» was strongly connected with the feature «Mountain» and the feature «Non-Tribal» with the features «Plains» or «Wet-rice culture». Down in the plains, or in the rare cases of terracing in the hills (the Apatanis among the Tani, the Angamis among the Naga), rice cultivation is done in two steps: the paddy shoots are grown in flooded gardens of small size called «nurseries»; then, they are transplanted into larger fields where they grow to maturity. This is linked to water in the big river and the monsoon. There is much work, especially for women who often do the transplanting. It produces big amounts of

rice, still more when fertilisers are used. As was mentioned above, the plains are densely inhabited and this leads to the notion of a higher level of civilization. Therefore, Savage or Tribal people who come from the hills and try to make a living in the plains are quickly civilized, and there is a general agreement that wet rice culture is something like a definition of decent life. This is one side of the truth.

Yet there are many Tribals in the plains, living on wet rice, who after some generations are civilised in some ways, but not in others. With the result that Assam, even in the restricted sense of the Assam Valley, is a more patchy and complicated mosaic (if this hackneyed metaphor is not misleading) than the average tourist would guess. Here again, language – in contrast to other features – gives interesting indications.

C/ Language distribution and historical population movements

Most populations have several names because (a) they have been given them by different outsider groups, (b) insider groups may have entertained various views about what they are, (c) several groups may become one. Even names like *Assam* or *Brahmaputra* are not protected against this all-pervading scepticism.

Present-day official names of «tribes» were coined by British administrators; only a few have been changed or added since then. Because official names were mainly devised in order to define distinct groups, each of them functions as a label. As far as Assam itself is concerned, British explorers and agents came mainly from northern Bengal and «contacted» tribal groups with the names that Bengali gave them. This is clearly illustrated in the reports written by Hamilton (Hamilton 1806), which were used in Martin's compilation (Martin 1830) and in official papers. The names we use in the next paragraphs still transmit this heritage – often a heavy burden.

It soon became clear that most «tribal» languages used in the Assam valley were close to each other, and two studies by Brandreth (Brandreth 1878) and Damant (Damant 1880) suggested, without a clear-cut demonstration, a list of connected languages in the plains. This group of languages is now often called «Boro-Garo», from the name of the two most famous representatives.

To-day (see Figure 6), the main Boro-Garo languages are Boro (Bor), Kokborok (Kok; officially: Tripuri), Garo (Gar), Rabha (Rab), Dimas (Dim), Tiwa (Tiw; officially: Lalung) and Deuri (or «Deori», Deu). Koch are not indicated on the map; some of them can be found where the letters «Rab» are pasted. Together, they number c. 3,000,000 speakers, most of them bilingual in Assamese or Bengali.

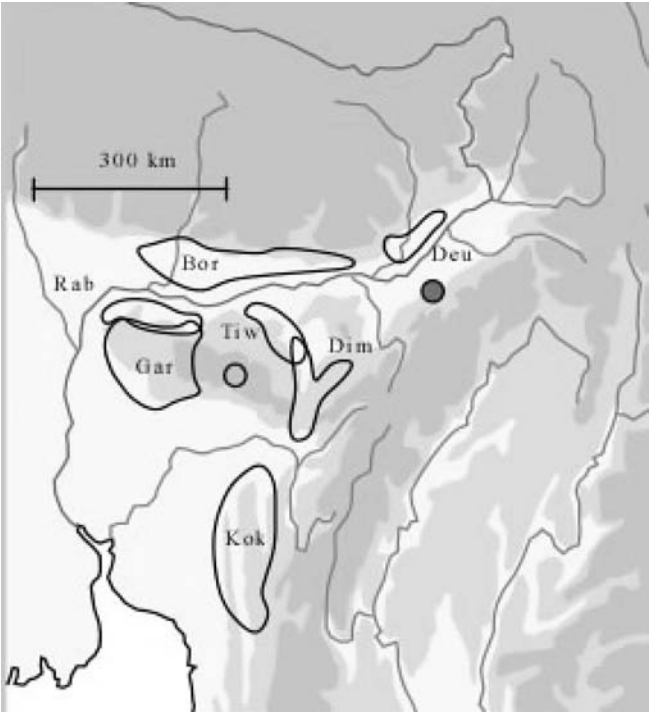


Figure 6. Location of the main Boro-Garo languages
Boro-Garo languages belong to the Tibeto-Burmese super-group. The red dot is for the Tai-Ahom group. The green dot is for the Mon-Khmer Khasi group

Boro	1222000
Tripuri	695000
Garo	676000
Rabha	139000
Dimasa	89000
Lalung	34000
Koch	26000
Deuri	18000
total	2899000

Figure 7. Number of speakers (Boro-Garo languages) in 2001
Source: Census of India.

The most important languages in the region, by number of speakers, are Bengali and Assamese, two closely related collections of dialects, both of them with a

literary tradition of several centuries, taught in schools and dominant in the media. They are «Indian languages» in the sense that they have common traits with languages of Northern India like Hindi, Gujarati, Nepali etc. especially with Oriya, the language of coastal Orissa. The success of these languages in Eastern, and North-eastern India is correlated with the colonization by Indian culture and people in various ways. Culture and people are not always identical because many people in Assam now speak Assamese, an Indian language (13,080,000 speakers in 2001), without being the descendants of Indian immigrants. Conversely, Indian immigrants have merged, more or less efficiently protected by their caste system (which is relatively loose in Assam), with the local «Tribals». It is precisely these not-so-clear characteristics and definitions that allow the lax redefinition of Tribals into Civilized via wet-rice production, and often their admission into Hinduism. Sometimes, as in the case of «Koch», new castes have been created to receive Tribal converts in the lap of Mother India.

However, Tribalism or something of the kind remains, along with Hinduism, in language, in the kinship system, and often also in food. We will skip the food topic here, because what we are interested in is not the criteria in themselves, but their shifting superimposition. Language is an efficient tool for analysis in this respect, since we can observe if people speak several languages, which of them, and why. Most people who can speak Boro-Garo languages also speak the dominant Assamese or Bengali languages.

Assamese is in some ways an adaptation of Indian modes of speech to Tribals. It was brought from central India by speakers that moved eastwards. Some of them are known because land grants written on copper plates are sometimes discovered; they never antedate the 5th century CE (Sharma 1978), but most of the present speakers are of mixed descent or not Indian at all by origin: they are simply local people whose ancestors shifted to Assamese and lost their earlier language(s). This process goes on, as the high level of bilingualism clearly shows. If Assamese and Bengali remain culturally dominant, and there is no sign of the contrary, these bilingual speakers will have children who will find it convenient to forget the smaller tribal languages. This is currently done in the Tripura State, where Kokborok is being silently and efficiently wiped out by Bengali. The remarkable fact is not this all too common picture of language shift but, on the contrary, the fact that so many languages in the plains where Assamese is dominant remain in use.

The black lines in Figure 6 that define areas for each language are a very clumsy representation since we already know that most of these speakers actually practise at least two languages; moreover, the areas criss-cross each other, sometimes they are densely populated with people speaking the language in question while sometimes the population is much more sparse. The Boro-Garo languages are not always contiguous: some interconnect or overlap, or not at all. The Kokborok area

in the south (it actually spreads into Myanmar and Bangladesh) seem cut off from the closest Boro-Garo language areas, but is actually smoothly connected to the Dimasa-speaking area.

The standard view for explaining this (see Jacquesson 2006, where more details are given) is the divergence process. It assumes there was originally only one language, or one dialect continuum (a more plausible idea in many cases), and that the speech area was fragmented by inroads or disrupted by over-extension. Figure 6 offers reasonable support to this view. The languages are spread out as if they were the contracted relics of earlier wet sponges that touched each other and might have formed only one corpus. The fact that the Deuri (or Deori) is isolated in Upper Assam can be explained from what we learn in local documents about the history of Upper Assam.

Boro-Garo languages were much spoken in Upper Assam in the past but the Ahom colonization from the 12th onwards, especially under the Ahom king Suhum in the 16th century, pushed the Dimasa speaking populations to the west and the south-west (where they now live) or absorbed them. The Ahom main point of settlement in Assam is indicated by the big red dot in Figure 6, and it is obvious that all the area around it is now devoid of any important Boro-Garo speaking group. Since all speakers of Ahom (a Tai language) had shifted to Assamese by the end of the 17th century, the result is an Assamese speaking area.

The same is true regarding the big green dot in eastern Meghalaya even if the historical reasons are just the opposite. This green dot is a symbol for the Mon-Khmer Khasi languages, still very much spoken in Eastern Meghalaya (912,000 speakers in 2001). It is very likely that these people had been speaking earlier forms of their Khasi languages for a long time. For instance, they are matrilinear and marriage is uxorilocal), a quite regular Mon-Khmer pattern, that we also find among the Garos, some Rabhas and some Tiwa people around them – all of them speaking Boro-Garo languages, which belong to the Tibeto-Burmese super-group, where matrilinearity is not common. This suggests that the Mon-Khmer influence was stronger in the past, or perhaps that some of these neighbouring people were in earlier times speaking some Mon-Khmer Khasi languages and later shifted to Boro-Garo languages but nevertheless retained their earlier social organization.

These two dots, the Mon-Khmer green one and the Tai red one, account for the two major areas where Boro-Garo languages are missing, and contribute to the idea that some time in the past these languages were spread all over the lower country.

D/ The logic of being somewhere

This first result derives from the evaluation (Jacquesson 2005) of the cohesiveness of the Boro-Garo linguistic unit and from the analysis of historical sources that can explain the present-day linguistic and/or ethnic maps (Jacquesson 2006). When we look into the details, two kinds of interesting perspectives appear. The first one is about «reticulation», the second one about «relation».

Figure 8 shows a simplified map of a rather standard situation in Assam. Other districts would require the indication of different groups but the overall result would be identical: ethnic groups usually do not define territories, at least not until recently. In Narayanpur circle, we find villages speaking Assamese, Tai-Khamti, Deuri (a Boro-Garo language) and Mishing (a Tani language). Normally, inside each village, only one of these languages is spoken (but one uses Assamese with visitors, or for specific situations like counting objects). The author studied this distribution in this circle, and knows when the different groups settled there. But our main point is the obvious fact that «ethnic spreading» is not always continuous: villages are cohesive (each village normally belongs to one group only) but the circle is not. The effect is, on the contrary, like bits and pieces distributed more or less at random. Colleagues studying other regions of Assam (in the Brahmaputra project, one follow-up of the «East meets West» ESF Project) reach similar conclusions with more refined analyses.

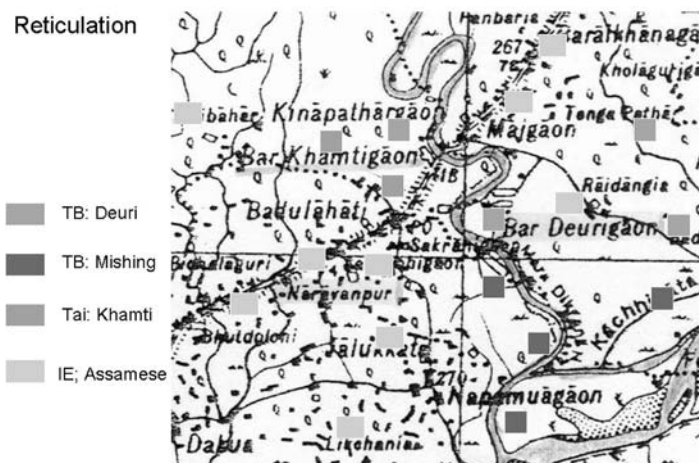


Figure 8. Linguistic reticulation (part of Lakhimpur district, Narayanpur circle). Each square is a village

In some respects, we are sent back to our earlier observations about dense and sparse, but now with a different background. It is true that some language groups are more dense in some places, less so in other ones, and quite absent elsewhere; but this distribution is usually complemented by the distribution of other language groups, so that the broad effect of a region being, for instance, rather dense does not tell you directly about the composition of this population, its linguistic (or ethnic, or political etc.) cohesiveness, and – most importantly, its possible reaction to upheaval or trauma. Indeed you often have to turn to «local knowledge» – often in the sense given to it by Clifford Geertz (Geertz 1983).

The consequence for historical interpretation is obvious. Unless you believe this patchy and mixed distribution is only a recent fashionable trend, you have to take it into account in historical «reconstructions», or views backwards. The first impression is that it blurs or obfuscates any possible view backwards: you cannot know the exact distribution of groups when you have no access to historical sources. The reticulation is out of reach, probably for ever. But what does not escape is the fact of the reticulation and the logic of social space.

Relation, our second topic in this section, is about the quality of the link between interrelated spots. When studying language history and appreciating language relatedness, phonological structure, prosody and grammar are as important as lexicon in the familiarity you feel with a language. Comparative linguistics is also concerned with the wide scope, not lexicon only, and is not restricted to quantitative approaches. The comparison of lexical data for evaluating a degree of relatedness is interesting in specific contexts, and usually useful between languages that are rather close to each other, but it never sums up a degree of relatedness. Technical comparison of Boro-Garo languages is interesting here: it gives insights into what is common life in the long run. The fabric of a Boro-Garo word is elusive because the Boro-Garo languages are distributed along a quasi-continuum. One end (Garo) has many consonants and few vowels, and the other end (Boro) has many vowels and few consonants. The standard way of «reconstructing» a proto-word in such cases is to pile up the features intelligently: to consider that the proto-word had many consonants *and* many vowels, and that some languages subsequently lost either this or that. This makes sense, because if words lose too much, they lose in distinction and become too much alike; therefore, it is only natural that some dialects or languages lose only in one respect, while others lose in other respects only.

The example of «to buy» (Figure 9) in some Boro-Garo languages exemplifies the continuum. The reconstructed form (marked with *, to show that it is not an actual form, only a useful hypothesis) is **brai* because some languages have *br-* and some have *-ai* (and Dimasa has *brai* indeed) but most have only one of them. The standard interpretation is that some have retained the *br-* but lost the *-ai* and conversely.

	buy	br-	-ai
	*brai		
Garó	bre-	+	-
Rabha	pri-	+	-
Tiwa	pre-	+	-
Dimasa	brai-	+	+
Boro	bai-	-	+
Kokborok	pai-	-	+
Deuri	zi-		

Figure 9. opposite tendencies in Boro-Garó languages

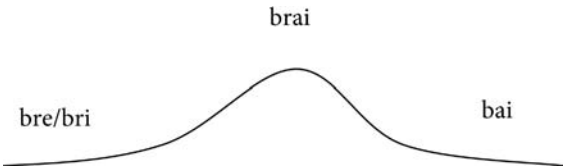


Figure 10. A readymade transposition from typology to space

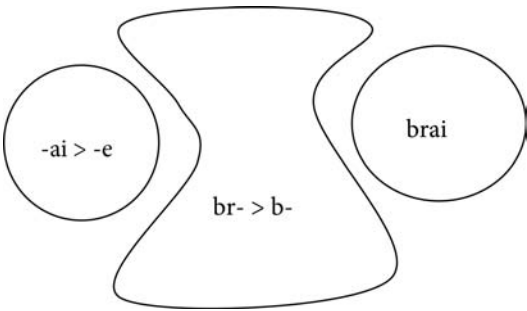


Figure 11. A plausible pattern for the Boro-Garó word fabric

If we suppose that the geographical distribution of speakers has not been troubled, or not so much, we may think of applying the continuum-like distribution on the map. The model would then be something like Figure 10.

It is very difficult to find a straightforward adaptation of the model to the geographical distribution of present-day languages. The complete form is typical of Dimasa and of some dialects related to Kokborok, and both languages are now

peripheral. The «bre» form is conveniently found only in Meghalaya languages. The «bai» form is found in Boro and Kokborok, the most distant languages in the group.

Yet, there is a possible explanation, if we admit that there were two distinct innovations rather than one organic continuum. This is suggested in Figure 11.

The drawing suggests that the older form *brai* was preserved, as often happens, only in peripheral languages, namely Dimasa, which we know (see above) to have been spoken in Upper Assam before the speakers were pushed westwards by the Ahom warfare. In Meghalaya and around, an innovation occurred that transformed the pronunciation of earlier diphthongs into simple vowels. In the major areas, north or south and right in the middle, a sweeping tendency to simplify the earlier consonant clusters occurred. Perhaps the speakers whose «ai» had shifted to a simple vowel resisted this because, as we described above, they could not oversimplify the pattern: this would imply the consequence that the «ai > e» shift would antedate the «br > b» shift.

What are the conclusions we can draw from such a pattern? We can conclude that the structure of a language only comes second when transformations occur, while the social links come first. What is important is which people communicate with each other because these communication tracks explain the spreading of an innovation. Here, the «br- > b-» spreading shows that communication was good enough among these people on the north and south axis, and indeed we know from historical sources that, in early British times, common taxes were still collected north and south for the common traditional (non-Ahom) power. The lack of «br- > b-» spreading shows that, when this change was innovative and spreading, those Dimasa speakers who were not influenced formed a relatively secluded community, either because there were few communications, or because the community was self-centred.

Conclusion: Local and wider knowledge about populations

All students of Assam have heard about the two kingdoms of pre-Ahom times, which the Chronicles call the Kachari and the Chutiya kingdoms. Before the British labelled the various populations of Assam, and still now in everyday speech, «Kachari» is intended first of all for Boro people. Later, in Endle times, a superdistinction was made between «Plains Kachari» meaning Boro, and «Hills Kachari» meaning Dimasa because Endle and others realized the closeness of those two groups in language and customs. In those times, Tripura was under a rajah, and moreover it was approached not from West Bengal, but by different officers, who worked in Sylhet and Chittagong.

If we take these conditions of contact (see section C) into account, and look further into the past, we may take as an interesting possibility that the Chronicles had something in view, and that earlier stages of the Dimasa community (under «Chutiya») were for some reason, maybe political, distinct enough from the earlier Boro + Kokborok stages (part of it under «Kachari»). This would fit into the picture of linguistic features described above.

Conversely, this and other descriptions given above show that the linguistic picture can be re-modelled for a number of reasons, as was the label «Tribal», or for that matter the label «Naga». The political map of earlier Assam, say before the Ahom conquest in the 16th century, is difficult to ascertain. But what is certain is that political frames, language links, economic settings, demography also (we skipped the disasters of wars), although often not contradictory, remain distinct as causes or parameters. It is then not reasonable to assess, as was done so often in the past that «peoples» are massive entities, walking across the landscapes with unchanging characteristics and identities.

References

- Bourdieu, Pierre. 1979. *La Distinction, critique sociale du jugement*. Paris: Editions de Minuit.
[trans. by R. Nice: *Distinction: a social critique of the judgement of taste*. Routledge & Kegan Paul, 1984.]
- Brahmaputra Project. www.vjf.cnrs.fr/brahmaputra
- Brandreth, E. J. 1878. "On the non-Aryan Languages of India." *Journal of the Royal Asiatic Society*, 10: 1–32.
- Damant, G. H. 1880. "Notes on the locality and population of the Tribes dwelling between the Brahmaputra and Ningthi rivers." *Journal of the Royal Asiatic Society*, 12: 228–258.
- Endle, Sidney. 1884. *Outline Grammar of the Kachari (Bara) Language*. Shillong: Assam Press.
- Geertz, Clifford. 1983. *Local Knowledge*. New York: Basic Books.
- Hamilton 1806. Manuscript documents, British Library, Indian Records.
- Jacquesson, François. 2005. *Le deuri, langue tibéto-birmane d'Assam*. Paris-Louvain: Peeters.
- Jacquesson, François. 2005. *Les langues bodo-garo (tibéto-birman d'Assam. Essai de linguistique historique et comparative*. Thèse d'Habilitation, Université Paris-3.
- Jacquesson, François. 2006. «La reconstruction linguistique du passé. Le cas des langues boro-garo.» *Bulletin de la Société de Linguistique de Paris* 101/1: 273–303.
- Martin, R. M. 1838. *The history, antiquities, topography and statistics of Eastern India, comprising the districts of Behar, Shahabad, Bhagulpoor, Goruckpoor, Dinajepoor, Puraniya, Ronggopoor, and Assam*. London: W.H. Allen.
- Sharma, M. M. 1978. *Inscriptions of Ancient Assam*. Guwahati: Gauhati University.

Genetic and linguistic diversity in Central Asia

Evelyne Heyer and Philippe Menneccier

In this study, we used genetic and linguistic data that we collected in Central Asia, to better understand how genetic and linguistic diversity correlates in a contact zone. We assessed the levels of genetic differentiation with mitochondrial, Y-chromosomal and autosomal data from 26 populations (1300 individuals) from the two major linguistic groups in Central Asia: Indo-European and Altaic. We computed the linguistic distance between populations with lexical data from several individuals per population.

Our results show that the genetic diversity in the area clearly clusters in two groups explained by the linguistic, one that includes the Indo-Iranian populations and the other one the Turkic populations except for Uzbek populations. Also, for two populations we have detected a shift in language that occurred likely through elite-dominance effect. Furthermore computing linguistic distances based on lexical data (Levenstein distance) we find a strong correlation between genetic distances and linguistic distances but no correlation between genetic and geographical distances.

In conclusion, Central Asia is an area where linguistic but not geography correlates with genetic diversity, highlighting the importance of a cultural trait in shaping genetic diversity in our species.

Introduction

Cavalli-Sforza's group (Cavalli-Sforza, Piazza, Menozzi, & Mountain, 1988) was first to attract the attention of scientific communities to the possible synchronization between linguistic and genetic evolution: he compared a tree summarizing the genetic differences of human populations on a global scale with the classification of world's languages — consisting of 17 families or phyla — suggested by Merrit Ruhlen (Ruhlen, 1987), a collaborator of Joseph Greenberg, and established a close match between such trees. Other researchers were reluctant to use genealogical trees either of languages or of populations. They developed another approach: they compared the results of genetic surveys, usually consisting of distance matrices and frequency vectors, with mathematically computed matrices of

linguistic diversity in order to statistically compute a correlation among genetic and linguistic variability (Sokal, 1988; Poloni et al., 1997; Lum, Cann, Martinson, & Jorde, 1998). This distance based method provides a statistical way to measure not only the correlation among genetic and linguistic distances but also their correlation with geographical distances. Using this method Rosser et al. (Rosser et al., 2000) argued that the correlation among genetic and linguistic distance was misleading. Being both highly correlated with geographic distance, they seemed inter-related but, when geographic distance was taken into account, their correlation disappears, a result that is in contradiction with previous studies based on other markers (Poloni et al., 1997; Sokal, 1988). In a more recent study on a worldwide analysis of human genetic diversity, Belle and Barbujani (Belle & Barbujani, 2007) show that genetic differences appear to more closely reflect geographic than linguistic differentiation. However, language differences also have a detectable effect on DNA diversity at the genomic level, above and beyond the effects of geographic distance. On a more regional scale, Lansing and co-workers (Lansing et al., 2007) reach a similar conclusion: when taking into account the geographical distances, linguistic and genetic distances still correlates. This marginal correlation could be explained by a reduced gene flow because of linguistic differences (Barbujani & Sokal, 1990) or could mean that languages and genes coevolved from a single common ancestor (i.e., were shaped by the same events in population history) (Nettle & Harriss, 2003).

In this scientific frame, we focus on Central Asia. In such area two major language families coexist: Indo-European languages and Altaic languages. The aim of our project was to compare genetic and linguistic diversity in order to better understand how a contact zone works.

Historical background

Located in the Eurasian heartland, Central Asia encompasses a vast territory limited to the east by the Pamir and Tien Shan mountain ranges, to the west by the Caspian Sea, to the north by the Russian taiga and to the south by the Iranian deserts and Afghan mountains. It comprises the ex-Soviet republics of Turkmenistan, Uzbekistan, Kyrgyzstan, Tajikistan and part of Kazakhstan. This area is crossed by the Silk Road.

Long term settlement

Archaeological remains testify the ancient peopling of Central Asia, eight hundred thousand years ago (Ranov & Schäfer, 2000). Evidence has also been uncovered of

settlement in the region during the Upper and Middle Palaeolithic Eras. This evidence includes Neandertal remains found in Teshik-Tash in Uzbekistan, the one of the most eastern remains (Okladnikov, 1939). First modern humans are documented before 30000BP (Trinkaus, 2005). Agriculture developed during the Neolithic Period, and a number of regional cultures developed endogenously in the region, with little input from external populations, with the notable exception of the Jeitun culture in Turkmenistan, which seems to have been imported from the Near East (Brunet, 1999). There seems to have been a mosaic of regional cultures, some of which had contacts with others, while others remained completely isolated. In the 1970 a series of excavation revealed the existence of a previously unknown Bronze Age civilization in Central Asia (Masson & Sarianidi, 1972). They revealed urban centers concentrated in southern Central Asia. This civilization variously labeled is most known as Oxus Civilization or Bactrian-Margiana-Archeological Complex (BMAC). An extensive craniometric study of human remains in the area “provides no support for models that suggest that Oxus Civilization inhabitants of the North Bactrian Oasis owe their origin to transplanted northern Iranian-affiliated populations from the urban centers of the Kopet Dagh foothill plain to the west or Indus Valley populations from the south-east“. The pattern of phenetic affinities rather suggest gene flow between northern Iranian-affiliated urban populations and an extant population within the north Bactrian oasis, emphasis the local temporal continuity of Central Asian populations (Hemphill, 1998). “The assertion that some local populations must have contributed to the formation of Oxus civilization population is compelling” (Hemphill, 1999). Further studies also show no relation between Burusho and BMAC (Willis & Hemphill, 2008). The first people in Central Asia for whom we have a historical record were the Scythians, around the 7th century B.C. Ancient Chinese and Herodoto texts describe these people as having European traits, and they were thought to speak Indo-Iranian languages.

Recent invasions

In the 2nd century B.C. and for more than 16 centuries, thereafter, the Chinese established a trade route (the Silk Road) across Eurasia, connecting the Mediterranean Basin to Eastern Asia. In the 3rd and 4th century A.D., multiple waves of Turkic invaders started to replace Indo-Iranians. Latter, the Turk-Mongol Empire became the largest empire of all time, from Mongolia to the Black Sea, following Genghis Khan's invasions in the 13th century A.D. Invasions occurred not only from the east, indeed Central Asia is the most eastern place reached by the Muslim invasion coming from Iran in the Xth century. The invaders spoke a West Iranian

language that swept through Central Asia and replaced the East Iranian languages except some “poche de résistance”, the most famous being the Yagnob.

All these movements of populations resulted in the presence of a considerable diversity of ethnic groups in Central Asia, most speaking Indo-Iranian or Turkic languages from the Indo-European and Altaic families of languages, respectively. These include Tajik sedentary agriculturists speaking an Indo-Iranian language and several Turkic populations speaking an Altaic language, who are traditionally nomadic, though some (e.g. Uzbeks) have shifted to a sedentary-agriculturist life style since the sixteenth century (Bregel, 2003).

The extent to which history and further western and eastern invasions have shaped the genetic and linguistic landscape in Central Asia is still unknown. Were there any language shifts, did some Indo-Iranian populations shift to Turkic language as a consequence of eastern invasion, or did some Turkic populations shift to Indo Iranian during the Muslim elite take-over?

In order to study this genetic and linguistic diversity we have designed a sampling strategy for both genetic and linguistic polymorphisms. For the genetic part of the study, we have analysed the uniparental markers: mitochondrial DNA that trace back the maternal history of the populations, whereas Y-chromosome gives insight into the paternal history of the populations and finally autosomal markers. Concerning the linguistic part of the study, we have recorded phonetic traits in most of the sampled populations.

Methodology

Sampled populations

We sampled 26 populations during 9 expeditions (1300 individuals) from West Uzbekistan to East Kyrgyzstan. These populations belong to seven ethnic groups (Tajiks, Yagnob, Kyrgyz, Karakalpaks, Uzbeks, Kazaks, and Turkmen) (see Table 1 and Figure 1 for information on the samples). The individuals sampled were representative of the ethnological diversity in the Central Asian steppes: Tajiks and Yagnobs are Indo-Iranian speakers (Indo-Iranian being a branch of the Indo-European language family), while Kazakhs, Turkmen, Karakalpaks, Kyrgyz and Uzbeks are Turkic speakers (Turkic being a branch of the Altaic language family).

For each sampled population, we have asked participants to give a small amount of blood (5ml) for DNA extraction. We carried out an ethnologic survey before sampling, during which individuals were asked about their recent genealogy, their geographical and ethnical origin and the ones of their four grand parents

Table 1. List of the sampled populations

Sampled populations (area) ^a	Acronym	Location	Long. ^b	Lat. ^b	Language family
Tajiks (Samarkand)	TJA	Uzbekistan / Tajikistan border	39.54	66.89	West Iranian
Tajiks (Samarkand)	TJU	Uzbekistan / Tajikistan border	39.50	67.27	West Iranian
Tajiks (Ferghana)	TJR	Uzbekistan / Kyrgyzstan border	40.36	71.28	West Iranian
Tajiks (Ferghana)	TJK	Uzbekistan / Kyrgyzstan border	40.25	71.87	West Iranian
Tajiks (Gharm)	TJE	Northern Tajikistan	39.12	70.67	West Iranian
Tajiks (Gharm)	TJN	Northern Tajikistan	38.09	68.81	West Iranian
Tajiks (Gharm)	TJT	Northern Tajikistan	39.11	70.86	West Iranian
Tajiks (Penjikent)	TDS	Tajikistan / Uzbekistan border	39.28	67.81	West Iranian
Tajiks (Penjikent)	TDU	Tajikistan / Uzbekistan border	39.44	68.26	West Iranian
Tajiks (Yagnobs from Dushanbe)	TJY	Middle Tajikistan	38.57	68.78	East Iranian
Uzbek (Ferghana)	UZA	Uzbekistan / Kyrgyzstan border	40.77	72.31	Turkic
Uzbek (Penjikent)	UZT	Northern Tajikistan	39.49	67.54	Turkic
Uzbek (Bukhara)	LUZn	Middle Uzbekistan	39.70	64.38	Turkic
Uzbek (Bukhara)	LUZa&k	Middle Uzbekistan	39.73	64.27	Turkic
Uzbek (Karakalpaks)	UZB	Western Uzbekistan	43.04	58.84	Turkic
Karakalpaks (Qongrat from Karakalpaks)	KKK	Western Uzbekistan	43.77	59.02	Turkic
Karakalpaks (On Tört Uruw from Karakalpaks)	OTU	Western Uzbekistan	42.94	59.78	Turkic
Kazaks (Karakalpaks)	KAZ	Western Uzbekistan	43.04	58.84	Turkic
Kazaks (Bukhara)	LKZ	Middle Uzbekistan	40.08	63.56	Turkic
Kyrgyz (Andijan)	KRA	Uzbekistan / Kyrgyzstan border	40.77	72.31	Turkic
Kyrgyz (Narin)	KRG	Eastern Kyrgyzstan	41.60	75.80	Turkic
Kyrgyz (Narin)	KRM	Eastern Kyrgyzstan	41.45	76.22	Turkic
Kyrgyz (Narin)	KRL	Eastern Kyrgyzstan	41.36	75.5	Turkic
Kyrgyz (Narin)	KRB	Eastern Kyrgyzstan	41.25	76	Turkic
Kyrgyz (Issyk Kul)	KRT	Eastern Kyrgyzstan	42.16	77.57	Turkic
Turkmen (Karakalpaks)	TUR	Western Uzbekistan	41.55	60.63	Turkic

^a Indo-Iranian populations are in white (Tajiks and Yagnob); Turkic are in grey (Kazaks, Karakalpaks, Kyrgyz, Turkmen and Uzbek).

^b Long., longitude; Lat., latitude.



Figure 1. Map of the area investigated. Note that the linguistic data of some locations are not yet included in the current dataset

and of their spouse. For each village, our sampling strategy was to sample individuals who were more distantly related than the first and second degree, and belonged to the same ethnic group. Such samples are considered as “populations” in our study. Informed consent was obtained from all participants

Molecular study

Total genomic DNA was isolated from blood samples by a standard phenol-chloroform extraction (Maniatis, Fritsh, & J., 1982).

Mt DNA

The first hypervariable segment (HVS-I) of the control region was sequenced in all samples, and variable positions were determined from position 16024 to 16383, as previously described (Quintana-Murci et al., 2004). The C-tract length variation at positions 16182 and 16183 in HVS-I was excluded from the analysis. Sequence quality was ensured as follows: each base pair was determined once with a forward and once with a reverse primer; any ambiguous base call was checked by additional

and independent PCR and sequencing reactions; all sequences were examined by two independent investigators.

Y-chromosome

Y chromosome diversity was assessed using a set of microsatellites, since these are variable in all populations and avoid the possible ascertainment bias associated with Y-SNPs. We typed 11 microsatellites on the Y chromosome. According to the protocol described by (Parkin et al., 2006), we have genotyped and analysed the microsatellites DYS19, DYS385 (DYS385a and DYS385b conjointly), DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426 and DYS439.

X and autosomal neutral markers

We used the informativeness for assignment index I_n (Rosenberg, Li, Ward, & Pritchard, 2003) to select subsets of short tandem repeats (STRs) on the X chromosome and the autosomes from the set of markers used in Rosenberg and colleagues' worldwide study (Rosenberg et al., 2002). This statistic measures the amount of information that multiallelic markers provide about individual ancestry. This index was calculated among a subset of 14 populations, chosen from the Rosenberg and colleagues' dataset to be genetically the closest to the Central Asian populations (Balochi, Brahui, Burusho, Hazara, Pathan, Shindi, Uygur, Han, Mongola, Yakut, Adygei, Russian, Druze and Palestinian). The rationale was to infer the information provided by individual loci about ancestry from this subset of populations, and to extrapolate the results to the populations studied here. For the X chromosome we chose 9 markers among the 11 with the highest I_n (Segurel et al., 2009). For autosomal data, we used the 'Screening Set10', which represented a total of 377 STRs, and chose 27 markers among the 30 with the highest I_n . (see (Martinez-Cruz et al., 2009) for more details).

Statistical analysis

In order to determine how overall genetic diversity is distributed within and between populations, an analysis of molecular variance (AMOVA) was performed using Arlequin v 2.0 software (Excoffier, Smouse, & Quattro, 1992). For mtDNA, the mutation model assumed was the Kimura 2-parameter model with a transversion/transition ratio of 10 and an alpha (Gamma shape parameter) of 0.26. For the Y-linked microsatellites, we used the RST genetic distance (Slatkin, 1995), which takes into account the probability of recurrent mutation. We performed a global AMOVA analysis including all populations and also considering several groupings corresponding to the linguistic and ethnic affiliation of populations. Correlations between genetic and geographical distances were performed with Mantel test implemented in R package (R Development Core Team, 2007).

Linguistic data.

Our first expedition was in the region of Karakalpakistan where speakers of Uzbek, Karakalpak and Kazakh language live. These 3 languages are very close and there is a mutual understanding between the different locutors in this area. In order to differentiate speakers of these languages, our first strategy was to choose 8 words that should emphasize the differences between the languages and to record the pronunciation of these 8 words from all sampled individuals. This first strategy correctly captured the differences between these 3 languages (Jacquesson, 2002) but was limited because it could not be extended to the whole Central Asian area, and the records of the pronunciation were strongly dependent on the individual who recorded it. Therefore, we switched to another strategy for the next samples; we decided to use the Swadesh list. This list is based on essential basic words forming the core of any language and is used classically by linguists. Altogether our data consist of a set of 180 spoken words for each of the 78 respondents, distributed across 23 testsites in the area. The words were recorded acoustically and transcribed in IPA which was subsequently recoded as X-SAMPA. The words are mostly members of the Swadesh-group of common basic words, 84 of them belonging to the Swadesh-100 group and 163 of them belonging to the Swadesh-200 group. The data were all only collected by Philippe Mennecier, the linguist of our team. We recorded 3 to 5 individuals per sampling area. Part of the data is still under analysis.

Linguistic distance method

We use a distance already developed for dialectometric research in the ALFA Informatica Groningen group. The software package L04, developed by Peter Kleiweg, was adopted for analysis (http://www.let.rug.nl/_kleiweg/L04). This package contains several methods to analyze phonological and lexical data statistically, building on the Levenshtein measure. The focus is on the comparison of pronunciation data such as IPA transcriptions.

Note that for each concept in the Swadesh list, and for each pair of sites we obtain a pronunciation distance—the distance between the pronunciation at the one site from the pronunciation at the other.

The first part of the study made in collaboration with the Groningen group was to test the methodology. We showed that the pronunciation distance techniques perform well in the preliminary classification of varieties even when the dataset includes unrelated varieties (Van der Ark, Mennecier, Nerbonne, & Manni, 2007).

Results and discussion on genetic diversity:

Autosomal markers: strong differences between the two main language families

The genetic diversity in the area clearly clusters in two groups, one that includes the Indo-Iranian populations and the other one the Turkic populations except for Uzbek (see Figure 2). Uzbek populations can be found both in the Indo-Iranian cluster and in the Turkic cluster, this is the subject of a specific paragraph further in the text. A formal test made by Amova on language family partition is highly significant ($p < 0.01$). When put in the broader perspective of Eurasian genetic diversity, we observe that the Turkic group is more closely related to eastern Eurasia, and the Indo-European is more closely related to Europe and Pakistan (data not shown).

When excluding the Uzbek, the genetic diversity among Turkic populations was slightly smaller than among Indo-Iranian populations $F_{st} = 0.008$ [0.006–0.010] and $F_{st} = 0.014$ [0.012–0.016] respectively.

We did not find any evidence for a correlation between genetic and geographic distances, either taking into account whole populations ($p = 0.18$) or dividing them by language family ($p = 0.92$ for Indo-Iranian speakers, $p = 0.45$ for Turkic excluding Uzbek).

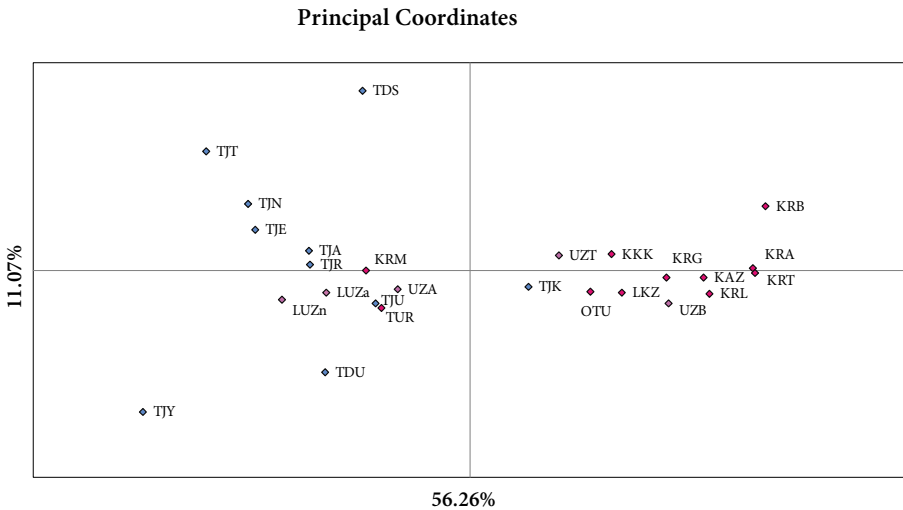


Figure 2. ACP of genetic distances among the populations. In red Turkic populations, in blue Indo-Iranian populations, in purple Uzbek populations

Social organisation: lesson from Uniparental markers and autosome –X comparison

These two groups of populations do not differ only by their language family but also have contrasted social organisation. One of the major finding of our research is to have evaluated the extent to which this social organisation shapes the genetic diversity focussing primarily on uniparental markers.

Although both Turkic nomadic and Indo-Iranian sedentary groups are patrilocally organised, only Turkic nomads retain a strong emphasis on their descent group affiliation. Indo-Iranian agriculturist societies are organised into families and marriage rules are based on kinship and geographical proximity with a strong preference for first-cousin marriages. Conversely, nomadic societies are organised into so-called “descent groups”, which include “lineages,” “clans,” and “tribes”. Individuals belonging to each of these descent groups claim common paternal ancestry. Individuals belong to the same lineage because they share a recent paternal common ancestor. Several lineages cluster in the same clan because they share a more distant common paternal ancestor, several clans cluster in one tribe because they share a more distant common ancestor. In addition, lineages and clans are exogamous entities, whereas tribes are endogamous. We have compared these ethnological genealogies with Y-chromosome data (Chaix et al., 2004). Indeed, we calculated a Y-chromosome Kinship coefficient and tested if individuals belonging to a given descent-group are closer paternally-related than individuals chosen at random. We have first shown that such claims are correct for individuals belonging to lineages and clans, but that links between individuals from a given tribe are cultural rather than biological (Chaix et al., 2004). Since this social organisation corresponds in part to a biological one (at the lineage and clan level), the next step was to measure the impact of such social organisation on genetic diversity.

To solve this question, we have compared the agriculturist societies with the traditionally nomadic societies. This social organisation has a strong impact on the intra population and interpopulation genetic diversity. We have shown that Turkic populations, but not Indo-Iranian populations, exhibited a significant loss of intra population genetic diversity for their Y chromosome (inherited from father to son), as a result of the dynamics of their patrilineal descent groups (Chaix et al., 2007). Regarding population differentiation we found that the overall genetic differentiation was much higher for the Y chromosome, as compared to mitochondrial DNA, among the 8 Turkic nomadic herder populations ($F_{ST}^{(Y)} = 0.177$ vs. $F_{ST}^{(mtDNA)} = 0.010$). In fact mitochondrial distances are rarely significant even when comparing populations from different ethnic group, whereas for the Y chromosome, there is always significant genetic differences among populations regardless of their ethnic group considered. For Indo-Iranian, the difference between Y chromosome differentiation and mitochondrial differentiation is less strong among

the 10 Indo-Iranian agriculturalist populations ($F_{ST}^{(Y)} = 0.069$ vs. $F_{ST}^{(mtDNA)} = 0.034$). Indo-Iranian populations are not only differentiated for their Y chromosome but also for their mitochondrial DNA (Chaix et al., 2007; Heyer et al., 2009).

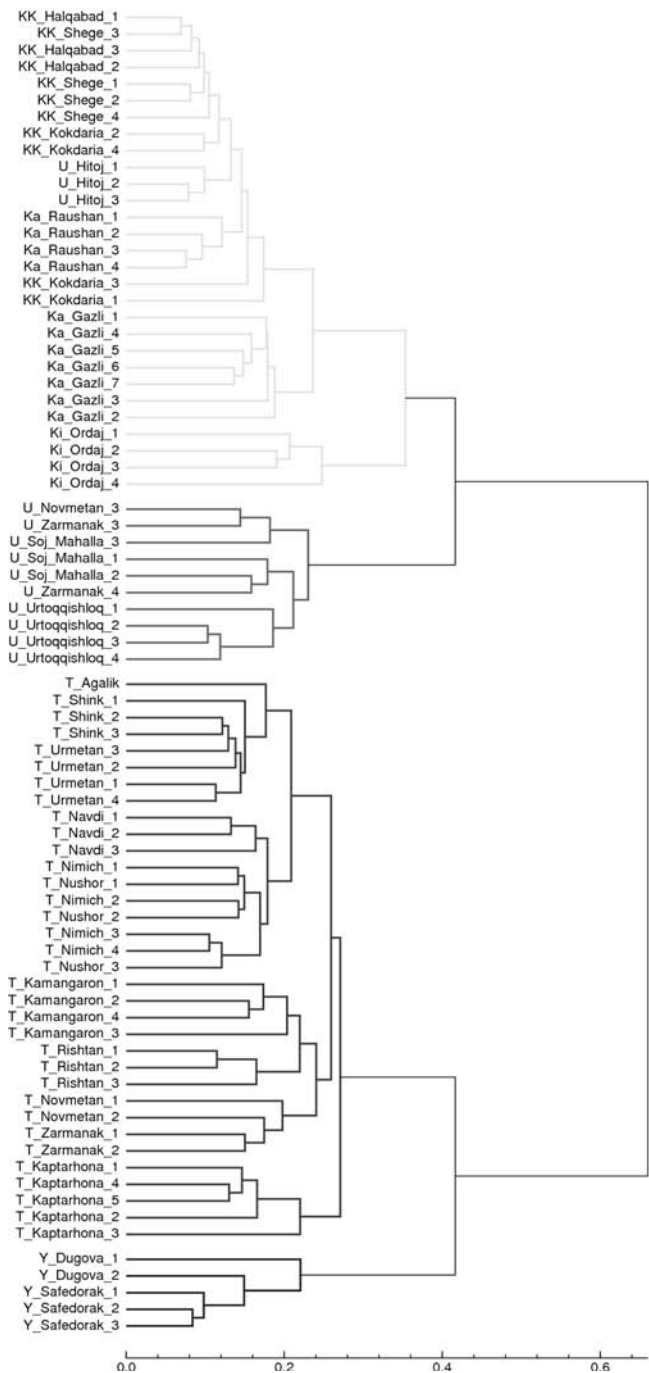
The fact that women are genetically less structured than men has been mainly explained by a higher migration rate of women, due to patrilocality (a tendency for men to stay in their birthplace while women move to their husband's house). Yet, since population differentiation depends upon the product of the effective number of individuals within each deme and the migration rate among demes, sex-biased dispersal and differences in male and female effective numbers have confounding effects on the comparison of genetic structure measured on uni-parentally inherited markers. We developed a new approach to analyze jointly autosomal and X-linked markers, to help understanding the sex-specific contributions to the genetic population differentiation. This method is based on the comparison of genetic distances measured on autosomal markers versus genetic distances measured on X-chromosome markers. In the absence of differences in male and female effective numbers, X chromosome distances are expected to be always higher than autosomal genetic distances. The exact relation between these two distances is:

$$F_{ST}^{(X)} \approx \frac{4F_{ST}^{(A)}}{4F_{ST}^{(A)} - 3(F_{ST}^{(A)} - 1)\left(\frac{1 + m_f/m}{2 - N_f/N}\right)}$$

Based on this relationship and varying the different parameters, we can estimate the ratio of male/female effective population size and male/female migration rate that are compatible with our data. With this new method, we showed that the contrast in genetic differentiation between Turkic and Indo-Iranians is not only the result of different pattern of migration of female, but also a sex specific difference in effective population size. We show that in these Turkic patrilineal herder groups of Central Asia, by contrast to Indo-Iranian bilineal agriculturalists, the effective number of women is higher than that of men. We interpret this result as the consequence of the patrilineal organization of herders, which tends to increase the relatedness among men (Segurel et al., 2008).

Results on linguistic diversity

Based on the linguistic distances, we built up a dendrogram (Fig3). We found the two major subdivisions between the two main language families: the first split show the dichotomy between Iranian languages versus Turk languages. Among the Iranian languages the split is clear between west-Iranian language (Tajik) and



KK = Karakalpaks; U = Uzbeks ; Ka = Kazaks ; Ki = Kyrgyz ; T = Tajiks ; Y = Yagnobs

Figure 3. UPGMA Dendrogram based on all words

east-Iranian language Yagnob that was spoken in the area before the Muslim invasion. Among Turk, the first split separates the Karluk language (Uzbek) from the others belonging to the Kiptchak languages. The last split is among the Kiptchak languages (Karakalpak, Kazakh, Kyrgyz). It has to be noted that Uzbek-speakers from Karakalpakistan retain their « original » Kipchak language.

The first result of our linguistic study was to extend the applicability of a range of computational techniques that have been implemented and successfully applied to problems in dialectology. The dialectological work has made prominent use of Levenstein distance (also known as edit distance) to analyze the relations among the various varieties of a range of European languages, including Irish, Dutch, Norwegian, Sardinian, German, American English, and Bulgarian. It distinguishes itself from the other dialectological research by analyzing data from more than one language group. Our work shows that this method can be further extend to compare languages belonging to different language families. It enables us to quantify intra and inter language family linguistic distances.

Correlation between genetic and linguistic distances

There is a good overall correlation between autosomal and linguistic data. The correlation is stronger with Y-chromosome than with mitochondrial data. When focusing inside the language family, this correlation still holds.

Both inside Indo-Iranian language family and inside Turkic language family, we find a significant genetic-linguistic correlation. This result was far from expected since in this area there is no genetic-geography correlation for autosomal data. The fact that the correlation is better with Y-chromosome than with mtDNA has already been described in previous study (Poloni et al., 1997)

Table 2. Genetic-linguistic correlations. Preliminary results on 16 populations

Correlation between genetic and linguistic distances						
Mantel tests						
Level	Indo-Iranian 10 populations		Turkic 6 populations		Global 16 populations	
	r	p	r	p	r	p
mtDNA-linguistic	0.3	0.18	0.6119	0.1218	0.08	0.2000
Y-linguistic	0.46	0.13	0.60755	0.0459	0.3	0.0059
Autosome-linguistic	0.45	0.04	0.77769	0.010988	0.54	0.00099

Cases where Linguistic and genetic do not match

The comparison between autosomal markers and linguistic data also enables us to identify several populations who have potentially gone through a language shift. Two Turkic populations, the TUR (Turkmen) and the KRM (Kyrgyz), cluster together with Indo-Iranians (Figure 2). Uzbeks populations can be found in the genetic cluster “Indo-European” and in the genetic cluster “Turkic”.

Turkmen: language shift through elite dominance?

The indigenous cultural history of the Turkmen in Turkmenistan can be dated back to 10,000 years B.C. and similarities between the cultures and technologies found in the archaeological record suggest that this region has been continually occupied since 6,000 B.C. Thus, the Turkmen, like the Tajiks (see above), may be the present-day descendants of populations established over long periods of time, accounting for Turkmen being genetically more similar to Tajiks than to Turkic (Figure 2). Subsequent Turk invasions may have led to the imposition of Turkic language with relatively little genetic admixture — perhaps through an elite dominance-driven linguistic replacement, as suggested for other nomad arrivals (Quintana-Murci et al., 2001). One mechanism for such an elite dominance is tribe assimilation. Turk invaders could have assimilated Indo-Iranian tribes. It is known that Turkmen consist of 2 major tribal unions – Teke and Iomud, the latter having incorporated some Iranian tribes. This could explain why these Turkic speakers have a gene pool closer to Indo-Iranian populations than to other Turkic.

Kyrgyz Several layers: Kyrgyz populations show a contrasted pattern of genetic affinities with other populations. KRM is genetically closer to Indo-Iranian populations than to other Turkic populations. Moreover, despite their geographic proximity, KRM and KRG are genetically very different. Indeed, the KRG population is genetically more similar to KRA and the other Turkic populations. This may be related to the origin of the Kyrgyz population, which has not been fully determined. The origin of the Kyrgyz living in Kyrgyzstan is still a matter of controversy in scholarly literature. Late in the eighth century the Kyrgyz state was a major rival of the Great Turkic Empire and later defeated the Uighur in the ninth century. The prevailing opinion now is that part of this Kyrgyz population moved from South Siberia to Kyrgyzstan in the fifteenth century and included some nomadic groups that inhabited the region for several centuries. The specific features of the KRM population may reflect a different layer in the history of Kyrgyz populations. Alternatively, the KRM population could be an Indo-Iranian population that has adopted a Turkic language.

The Uzbeks an heterogeneous group. Another intriguing case is the Uzbek case. Uzbeks were nomadic herders until they adopted a sedentary lifestyle and

became farmers toward the 16th Century. Soucek (Soucek, 2000) records that what is now called 'Uzbek' encompasses the seventeenth century Uzbek and former Chagatai Turk groups who were already settled in Uzbekistan. Therefore the name refers to a tribal union of different tribes including Chagatai Turks who were strongly mixed with Iranian dwellers of Central Asia. Our linguistic analysis clearly shows the existence of these two groups of Uzbek. The Uzbek from Hitoj in Karakalpakistan cluster with the Kiptchak group that includes the other Nomadic populations: Kazakh, Kyrgyzh, Karakalpak. On the other hand the other Uzbek populations form a different cluster: the Karluk group. This difference in linguistic can also be seen in the genetic data. The Uzbek from Karakalpakistan (label UZB) cluster with the Turkic group whereas several other Uzbek populations cluster in the Indo-Iranian group. Furthermore, parts of the 'Uzbek' individuals that we sampled in the area of Bukhara (LUZ samples) speak both Tajik and Uzbek at home.

Conclusion

When comparing linguistic and genetic data at a family language level, we find the classical results: linguistic and genetic differences correlate. But in Central Asia, this correlation cannot be explained by geographic distance since pairwise genetic distances between populations do not correlate with corresponding geographic ones. This means that there is a limited gene flow between populations belonging to the two different language groups irrespectful to the geographic distance between such populations. This is perhaps strengthened by the fact that the two language groups have contrasted social organisations. These different social organisations could also act as a social barrier reducing gene flow between the two groups.

Regarding the genetic data the populations clearly cluster in two groups which correspond to the two language family. This result enables us to identify some populations where linguistic and genetic do not match. Surprisingly, despite documented warrior's invasions, known for their ferocity (Attila, Genghis Khan) we find few language shifts which could have resulted from elite dominance.

Using phonetic distances we have compared linguistic and genetic diversity within each language family. We have classical results, meaning a strong correlation between linguistic and genetic distances. This result is quite intriguing for the Indo-Iranian group since all our sampled populations except one (the Yagnob) have shifted from East-Iranian languages to West-Iranian languages in the Xth century following Muslim invasion.

In this paper we presented preliminary results, research are still going on in order to explain our results and more exhaustive papers are planned in the near future. In particular, new currently developed methods in population genetics

studies should be soon available in order to date the divergence between populations in models which include migration between populations. This will enable us to better understand the underlying processes that result in a correlation between linguistic and genetic distances in the absence of any geographical correlation.

List of collaborators on our project on Central Asia:

In France:

Dr. R Chaix, geneticist, CNRS
Dr. Franz Manni, geneticist & linguist, MNHN
PhD Student Laure Ségurel
Dr. Renaud Vitalis, geneticist, CNRS
Dr. Begona Martinez Cruz, geneticist, CNRS
Dr. Lluís Quintana, geneticist, CNRS, Inst. Pasteur
Dr. Frederic Austerlitz, geneticist, CNRS
Dr. François Jacquesson, linguist, CNRS
Dr. Svetlana Jacquesson, anthropologist, IFEAC
Ing. Myriam Georges, Tech, CNRS

In Central Asia:

Dr. Tatiana Hegay, geneticist, Tashkent
Pr. Ruslan Ruzibakiev, geneticist, Tashkent
Dr. Almaz Aldashev, geneticist, Bishkek
Pr. Vadim Yagodin, archaeologist, Nukus
Pr. Firuza Nasyrova, geneticist, Douchanbé

Other countries

Dr Michael Hammer (USA)
Pr Mark Jobling (UK)
Dr. P Balaesque (UK)
Dr J Nerbonne (NL)

References

- Barbujani, G., & Sokal, R. R. (1990). Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc Natl Acad Sci U S A*, 87(5), 1816–1819.
- Belle, E. M., & Barbujani, G. (2007). Worldwide analysis of multiple microsatellites: language diversity has a detectable influence on DNA diversity. *Am J Phys Anthropol*, 133(4), 1137–1146.
- Bregel, Y. (2003). *An Historical Atlas of Central Asia*. Leiden, Boston: Brill.
- Brunet, F. (1999). La Néolithisation en Asie Centrale: un état de la question. *Paléorient*, 24, 27–48.

- Cavalli-Sforza, L. L., Piazza, A., Menozzi, P., & Mountain, J. (1988). Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc Natl Acad Sci U S A*, 85(16), 6002–6006.
- Chaix, R., Austerlitz, F., Khegay, T., Jacquesson, S., Hammer, M. F., Heyer, E., et al. (2004). The genetic or mythical ancestry of descent groups: lessons from the Y chromosome. *Am J Hum Genet*, 75(6), 1113–1116.
- Chaix, R., Quintana-Murci, L., Hegay, T., Hammer, M. F., Mobasher, Z., Austerlitz, F., et al. (2007). From social to genetic structures in central Asia. *Curr Biol*, 17(1), 43–48.
- Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131(2), 479–491.
- Hemphill, B. E. (1998). Biological affinities and adaptations of Bronze Age Bactrians: III. An initial craniometric assessment. *Am J Phys Anthropol*, 106(3), 329–348.
- Hemphill, B. E. (1999). Biological affinities and adaptations of Bronze Age Bactrians: IV. A craniometric investigation of Bactrian origins. *Am J Phys Anthropol*, 108(2), 173–192.
- Heyer, E., Balaesque, P., Jobling, M., Quintana-Murci, L., Chaix, R., Segurel, L., et al. (2009). Genetic diversity and the emergence of ethnic groups in Central Asia. In press. *BMC Genetics*.
- Jacquesson, F. (2002). Karakalpak et autres gens de l'Aral. In EDISUD (Ed.), *Cahiers d'Asie Centrale* (pp. 192). Tachkent.
- Lansing, J. S., Cox, M. P., Downey, S. S., Gabler, B. M., Hallmark, B., Karafet, T. M., et al. (2007). Coevolution of languages and genes on the island of Sumba, eastern Indonesia. *Proc Natl Acad Sci U S A*, 104(41), 16022–16026.
- Lum, J. K., Cann, R. L., Martinson, J. J., & Jorde, L. B. (1998). Mitochondrial and nuclear genetic relationships among Pacific Island and Asian populations. *Am J Hum Genet*, 63(2), 613–624.
- Maniatis, T., Fritsch, E. F., & J., S. (1982). Molecular cloning. A laboratory manual. New York: Cold Spring Laboratory.
- Martinez-Cruz, B., Vitalis, R., Austerlitz, F., Quintana-Murci, L., Georges, M., Hegay, T., et al. (2009). A multilocus genetic landscape in Central Asian human populations. *Submitted*.
- Masson, V. M., & Sarianidi, V. I. (1972). *Central Asia; Turkmenia before the Achaemenids*. [London]: Thames and Hudson.
- Nettle, D., & Harriss, L. (2003). Genetic and linguistic affinities between human populations in Eurasia and West Africa. *Hum Biol*, 75(3), 331–344.
- Okladnikov, A. (1939). Nakhodka Neandertal'tsa v Uzbekistane. *Vestnik Drevnei Istorii*, 1, 256–257.
- Parkin, E. J., Kraayenbrink, T., van Driem, G. L., Tshering Of Gaselo, K., de Knijff, P., & Jobling, M. A. (2006). 26-Locus Y-STR typing in a Bhutanese population sample. *Forensic Sci Int*, 161(1), 1–7.
- Poloni, E. S., Semino, O., Passarino, G., Santachiara-Benerecetti, A. S., Dupanloup, I., Langaney, A., et al. (1997). Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *Am J Hum Genet*, 61(5), 1015–1035.
- Quintana-Murci, L., Chaix, R., Wells, R. S., Behar, D. M., Sayar, H., Scozzari, R., et al. (2004). Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet*, 74(5), 827–845.
- Quintana-Murci, L., Krausz, C., Zerjal, T., Sayar, S. H., Hammer, M. F., Mehdi, S. Q., et al. (2001). Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am J Hum Genet*, 68(2), 537–542.

- R Development Core Team. (2007). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Ranov, V., & Schäfer, J. (2000). The Palaeolithic of the late Middle Pleistocene in Central Asia, 400–100 ka ago. In W.-E. M. Ronen A (Ed.), *Toward Modern Humans The Yabrudian and The Micoquian, 400–50 k-years ago* (pp. 77–93): Congress University of Haifa.
- Rosenberg, N. A., Li, L. M., Ward, R., & Pritchard, J. K. (2003). Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet*, 73(6), 1402–1422.
- Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., et al. (2002). Genetic structure of human populations. *Science*, 298(5602), 2381–2385.
- Rosser, Z. H., Zerjal, T., Hurles, M. E., Adojaan, M., Alavantic, D., Amorim, A., et al. (2000). Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet*, 67(6), 1526–1543.
- Ruhlen, M. (1987). *A guide to the world's languages*. Stanford, Calif.: Stanford University Press.
- Segurel, L., Martinez-Cruz, B., Quintana-Murci, L., Balaesque, P., Georges, M., Hegay, T., et al. (2008). Sex-specific genetic structure and Social Organization in Central Asia: Insights from a Multi-Locus Study. *PlosGenetics*.
- Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 139(1), 457–462.
- Sokal, R. R. (1988). Genetic, geographic, and linguistic distances in Europe. *Proc Natl Acad Sci U S A*, 85(5), 1722–1726.
- Soucek, S. (2000). *A history of inner Asia*. Cambridge; New York: Cambridge University Press.
- Trinkaus, E. (2005). Early modern humans. *Ann Rev Anthropol*, 34, 207–230.
- Van der Ark, R., Menecier, P., Nerbonne, J., & Manni, F. (2007). *Preliminary Identification of Language Groups and Loan Words in Central Asia*. Paper presented at the Proceedings of the RANLP Workshop on Computational Phonology Workshop at the conference Recent Advances in Natural Language Phonology Borovetz.
- Willis, C., & Hemphill, B. (2008). Are the Burusho an indigenous population of the Northern Areas, Pakistan? An odontometric investigation. *American Journal of Physical Anthropology*, 135(S46).

Genetic and linguistic borders in the Himalayan Region

Thirsa Kraaijenbrink¹, Emma J. Parkin²,
Denise R. Carvalho-Silva³, George L. van Driem¹,
Guido Barbujani⁴, Chris Tyler-Smith³, Mark A. Jobling²,
Peter de Knijff^{1,5}

There are a number of competing theories about the origins of the Himalayan peoples. These theories are largely based on linguistic and/or archaeological findings, sometimes supported by the results of small-scale genetic studies. A large-scale, ethnolinguistically-informed genetic study of the greater Himalayan region might provide a definitive model for historical population events in this region, and that is why the current study was undertaken.

The geographical area of the present-day states of Nepal and Bhutan could have served as ancient corridors for human migration through the Himalayas despite their geographical position immediately south of the highest land barrier. The findings also raise the question as to whether the southern slopes of the Himalayas could have harboured a myriad of refuge areas for the ancestral Tibeto-Burman population(s) during the last glacial maximum. Alternatively, if the multitude and diversity of language communities found in these countries is a reliable indication, they could be an ancient source of genetically differentiated populations and languages. A detailed genetic study of the Himalayan region, therefore, may not only provide insights into the uniqueness and antiquity of its residents, but may also shed light on the peopling of the Himalayas and eastern Asia in general.

-
1. Department of Human and Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands.
 2. Department of Genetics, University of Leicester, United Kingdom.
 3. The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom.
 4. Dipartimento di Biologia, Università di Ferrara, Italy.
 5. Corresponding author: P. de Knijff, Department of Human and Clinical Genetics, Leiden University Medical Centre, PO Box 9600, 2300 RC Leiden, The Netherlands. Phone + 31 71 526 9540, e-mail knijff@lumc.nl.

Using genetic data from 15 autosomal Short Tandem Repeat (STR) loci, we provide evidence that there is clear congruence between language and genetics. Populations speaking a language belonging to the Tibeto-Burman language family are genetically more similar to each other than to populations speaking a language belonging to the Indo-European language family. On the basis of language differences we can draw a linguistic boundary roughly running from east to west, just south of the border between India and Bhutan and through southern Nepal. A genetic boundary can be reconstructed along nearly the same route. We conclude from all these analyses that the populations of Nepal and Bhutan are likely to have originated outside their current locations, in regions where their language families are spoken, but need further work to suggest more precise origins.

Introduction

Isolation is crucial to both biological and linguistic change. Populations that are separated by barriers tend to diverge genetically because of genetic drift, and to undergo independent linguistic change, resulting in often parallel patterns of genetic and cultural differentiation. Geographical as well as cultural barriers reduce population contacts, thereby potentially increasing isolation between populations. However, both biological and linguistic change are influenced by the size of the population. It is thus also important to infer reliable information on past human demography.

The greater Himalayan region is ethnolinguistically the most complex area of the Eurasian continent. This region includes the highest land barrier on the face of this planet, and linguistic evidence unambiguously indicates that topography has shaped and channelled prehistoric population movements. The intricate ethnolinguistic mosaic of this region holds many keys to the peopling of the Eurasian continent as a whole. Whereas most language communities in the Himalayan region belong either to the Tibeto-Burman or Indo-European family, there are also Austroasiatic, Dravidian, Daic and Altaic language communities settled in the mountains, foothills and periphery of the Himalayas. Moreover, there are two language isolates, Burushaski and Kusunda, in the region. Linguistically, the Himalayas are sometimes thought to form the border between the Indo-European and Tibeto-Burman language families, though in fact the real linguistic border roughly runs parallel to the range through the hills and lowlands to the south (van Driem 2001). Some genetic studies have indicated the presence of a genetic barrier in this area, but these studies have mainly included population samples from China and India and not from populations within the Himalayan heartland: Nepal and Bhutan (Cordaux et al. 2004, Metspalu et al. 2004, Xue et al. 2006).

A very few studies include some Himalayan population samples (Cavalli-Sforza et al. 1994, Gayden et al. 2007) but were unable to sample extensively in this area.

The geographical area of the present-day states of Nepal and Bhutan could have been corridors for human migration through the Himalayas in ancient times despite their geographical position immediately south of the highest land barrier, the Himalayan mountain range: for people adapted to life at this altitude, they provide the most inviting localities. Or they could be seen as regions where human existence is difficult, and inhabited late in prehistory. Alternatively, if the multitude of language communities found in these countries is a reliable indication, they could be an ancient source of genetically differentiated populations and languages, a possible consequence of subdivision and extreme isolation over long periods. A detailed genetic study of the Himalayan region, therefore, may not only provide evidence for the uniqueness and antiquity of its residents, but may also shed light on the peopling of the Himalayas and eastern Asia in general.

In order to be able to analyse the possible correlation between the complex linguistic relationships and the genetic affinities among the many Himalayan populations and those of their neighbouring regions, we embarked upon two sampling expeditions to Nepal and Bhutan with the aim of providing answers to three major questions:

- Is there a correlation between language, genes and geography in the Himalayan region?
- Can we determine the genetic relationships (ancient ancestors) of the Nepalese and Bhutanese and deduce possible migration routes?
- Can we say something about relative ages of the various groups now living there, identifying and comparing “aboriginal” groups with the others?

In this article we will provide details of the expeditions, the samples, and the genetic systems tested. We will also describe in some detail the first autosomal DNA results. The analyses of mtDNA and Y-chromosomal data are not yet completed and will be described in a future publication.

Methods

Planning of the project

Initially, the aim of the study was to organise three expeditions, to Nepal, Bhutan and North and north-eastern India (specifically: Assam, Sikkim, and Arunachal Pradesh), to collect blood from the main ethnolinguistic groups of the greater Himalayan region. Unfortunately, it eventually turned out to be impossible

– within the time frame of our funding – to collect samples in India. Therefore, the project was restricted to the analyses of the Nepalese and Bhutanese ethnolinguistic groups, as described in detail below.

We organised two major expeditions, one to Nepal and one to Bhutan. The first expedition was aimed at collecting blood samples from Nepal's populations. During this expedition, held in December 2002 and January 2003, the team was assisted by several Nepali assistants from various language communities. The work in Nepal was carried out with the knowledge and cooperation of representatives of local groups and Tribhuvan University at Kirtipur. With the valuable assistance and guidance of Prof. Dr. Nirmal Man Tuladhar (Professor of Linguistics at Tribhuvan University), representatives of the ethnolinguistic groups were contacted and asked for cooperation in the project. The names of these representatives can be found in the detailed acknowledgements.

Blood donation was on a voluntary basis, often after discussing the project with local (language) communities, and a blood-sample (10 ml) was only taken if the donor had read, understood and signed the informed consent form. When a donor was unable to read or write, the consent text was read to the donor in his/her local language, after which one of the project's co-workers filled in the donor's data on the form. For some communities, detailed explanation in the local language was given and video-recorded for archival purposes.

To ascertain that a person belonged to a certain ethnolinguistic group or caste, the donor's name and place of birth were systematically checked against what is known about the names adopted by members of Nepal's diverse ethnic groups and the geographical spread of the group with which the person identified himself or herself. In addition, several team members and most of the project's co-workers speak one or more relevant Nepali languages. To consider that a person was not admixed, both parents had to belong to the same group.

The second expedition, which was aimed at collecting blood samples from Bhutanese populations, was completed in October and November 2003. The work in Bhutan was carried out with the knowledge, permission and cooperation of the Royal Government of Bhutan at Thimphu. As in Nepal, blood donation was on a voluntary basis and sampling of 10 ml of blood was only carried out after the donor had read, understood and signed the consent form. Again, when a donor was unable to read or write, the consent text was read to the donor and his or her data were recorded by one of the project's co-workers.

Donors had been pre-selected by representatives of the Royal Government of Bhutan, based on the same criteria as were used in Nepal (the only difference being that not all ethnolinguistic groups in Bhutan use group-specific names).

Blood was collected throughout Bhutan during four field trips. Members of some major Bhutanese groups were sampled in and around Thimphu (the capital

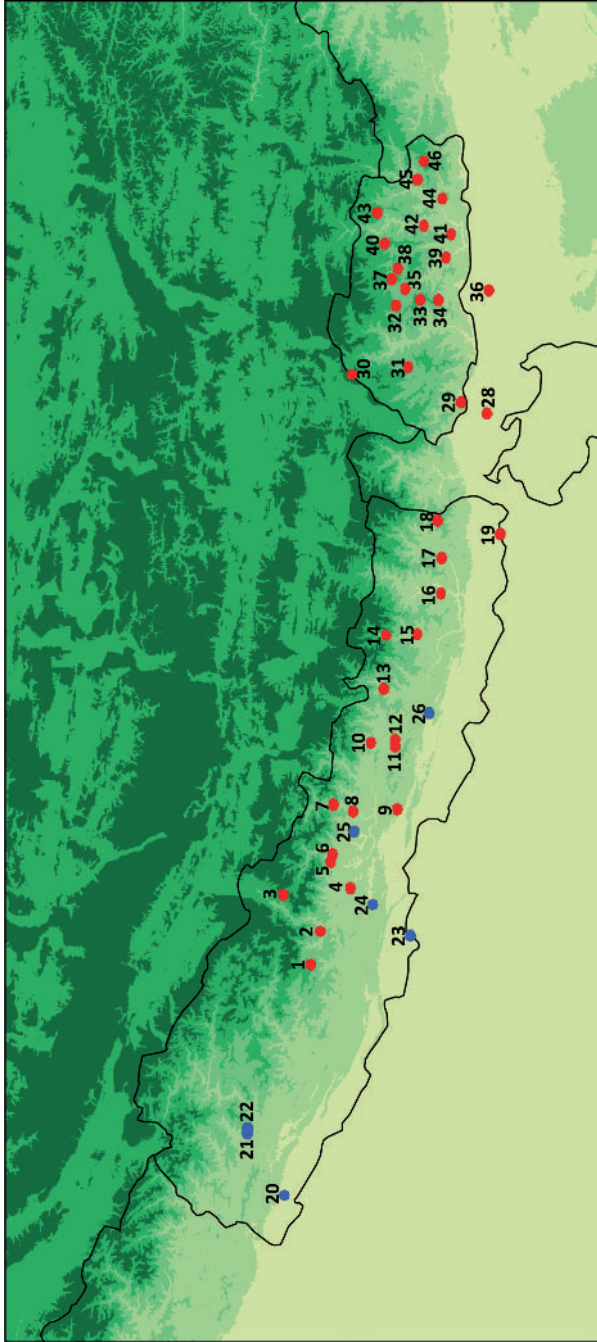


Figure 1. Distribution of ethnolinguistic groups sampled in Nepal and Bhutan

In Nepal, the blue dots reflect the Indo-European language group centres and the red dots reflect language group centres of the Tibeto-Burman speaking populations. In Bhutan, all populations speak a Tibeto-Burman language. Numbers (see also below) correspond with the numbers in Table 1 (p. 198–201): 1, Kham; 2, Chantyal; 3, Thakali; 4, Magar; 5, Gurung; 6, Dura; 7, Ghale; 8, Barâm; 9, Chepang; 10, Tamang; 11, Newar; 12, High Caste Newar; 13, Thangmi; 14, Sherpa; 15, Western Kiranti; 16, Central Kiranti; 17, Eastern Kiranti; 18, Limbu; 19, Dhimal; 20, Indo European; 21, Bahun; 22, Chettri; 23, Tharu; 24, Majhi; 25, Kumal; 26, Indo European / Tibeto Burman substrate; 28, Toto; 29, Lhopku; 30, Layap; 31, 'Ngalop; 32, Lakha; 33, Mangde; 34, Black Mountain Mõnpa; 35, Nup; 36, Bodo; 37, Bumthang; 38, Brokkat; 39, Khengpa; 40, Kurtöp; 41, Gongduk; 42, Chali; 43, Dzala; 44, Tshangla; 45, Dakpa; 46, Brokpa.

city) and when encountered during any of the four expeditions. During the various field trips in Bhutan, we were also able to collect samples from two Tibeto-Burman-speaking populations from northern India: the Bodo and the Toto. Table 1 (p. 198–201) presents descriptive statistics of the sampled individuals. Figure 1 illustrates the geographical coordinates of the centres of the various language groups sampled by us and included in our study.

Sample processing

After blood sampling, a small amount of blood was spotted in duplicate onto FTA filter-paper cards for archival purposes (4 spots of ~1 cm diameter per FTA card).

The blood samples were sent to the Netherlands via DHL as soon as possible after sampling (tubes and FTA cards were sent separately). One set of FTA cards is currently stored in Leiden and one set in Leicester for future reference.

The blood in the tubes was used for DNA isolation, using the Autopure LS® from Gentra Systems, according to the manufacturer's specifications. All blood samples yielded a sufficient amount of good quality DNA. Aliquots of all samples were shipped to the United Kingdom for Y-chromosomal and mitochondrial-DNA analyses.

Genotyping

After DNA isolation, all Nepalese and Bhutanese samples were genotyped for 21 forensic autosomal Short Tandem Repeat (STR) loci, contained in three commercially available kits: Powerplex 16 (Promega), AMPFISTR Identifier (Applied Biosystems) and FFFL (Promega).

To our own data, we added data from many reference populations from India and China (Kraaijenbrink et al. in prep., Rajkumar and Kashyap 2003, Gaikwad and Kashyap 2003, Neeta and Kashyap 2004, Kasyap et al. 2004, Hima Bindu et al. 2005, Krithika et al. 2006, Kasyap et al. 2006, Hima Bindu et al. 2007, Xue et al. 2006, Quintana-Murci et al. 2001, Lee et al. 2004). From many of these reference populations, only genotypes of the loci contained in the Powerplex 16 kit were available. Therefore it was decided to limit the analyses reported here to the 15 autosomal STR loci contained in this kit.

Statistical analyses

Population structure was examined using the program Structure 2.2 (Pritchard et al. 2000, Falush et al. 2003, Falush et al. 2007) based on the admixture model with correlation between allele frequencies across clusters. For each number of clusters

K, five independent Structure runs were performed, all using a burn-in of 20,000 iterations, followed by 10,000 iterations of MCMC for estimates of clustering.

Pairwise F_{ST} for all population pairs was calculated using the Excel add-in Genalex 6.1 (Peakall and Smouse, 2006). In order to compare with the results obtained using Structure, the pairwise F_{ST} values were used in multi-dimensional scaling (MDS) analyses performed with the program NCSS. The first two dimensions resulting from the NCSS analyses were used for creating an MDS plot in Excel.

From the Structure analyses with $K=2$, we took the estimated values of attribution to the two clusters of each of the population samples as input for a spatial distribution map using the Kriging procedure with the Surfer 8 software (Golden Software, <http://www.goldensoftware.com>).

Results and discussion

In total, we collected DNA samples from 947 unrelated Nepalese volunteers (764 males and 183 females) and 1029 unrelated Bhutanese volunteers (839 males and 190 females), belonging to 40 major ethnolinguistic groups from the Tibeto-Burman family, and 11 ethnolinguistic groups from the Indo-European family (Table 1 p. 198–201). All samples were analysed for the 21 forensic autosomal STRs comprised in the Powerplex16, Identifiler and FFFL kits. Due to small sample size for some population samples, a total of 23 individuals from Nepal (see Table 1 p. 198–201) were not included in this study.

The general allele frequency distributions of these autosomal STR loci for Nepal and Bhutan have been published previously (Kraaijenbrink et al. 2007a, Kraaijenbrink et al. 2007b). When analysing the data in Structure, the two-cluster model ($K=2$) was the best fit of our total dataset. All runs for $K=2$ produced the distribution shown in Figure 2A. Most populations belonging to the Indo-European family are grouped together in one cluster which is predominantly blue in Figure 2A, and most populations belonging to the Tibeto-Burman family are grouped together in the other cluster which is predominantly yellow in Figure 2A.

When we increased the number of clusters, the clear distinction between the Tibeto-Burman and the Indo-European language groups was lost. Instead, with $K=6$, three populations became clearly clustered individually (Figure 2B). These populations, the Toto from North India (T in Figure 2B), and the Lhokpu (L in Figure 2B) and the Black Mountain Mönpa (M in Figure 2B) from Bhutan, are known to have been almost completely isolated from their neighbouring populations until relatively recently, due to both geographical and cultural barriers, which provides a possible explanation for this separate clustering.

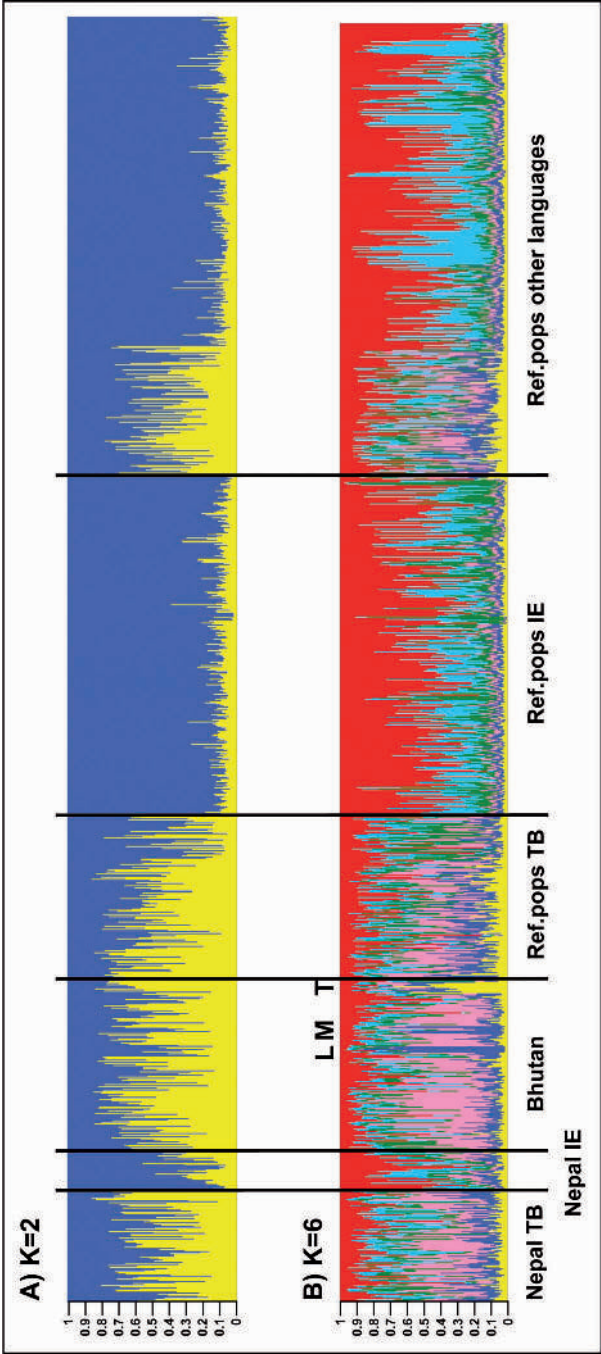


Figure 2. Results of unsupervised Structure analyses

The colours represent the proportion of inferred ancestry from K ancestral populations. For K=2 (2A), the inferred ancestry largely corresponds to the linguistic family to which the populations belong with Tibeto-Burman speaking populations mostly being assigned to the yellow cluster and Indo-European speaking populations to the blue cluster. For K=6 (2B) the majority of the “linguistic clustering” is lost in favour of the separation of the Lhokpu (L), Black Mountain Mönpa (M) and Toto (T), although some differences can still be observed between the Tibeto-Burman and Indo-European populations.

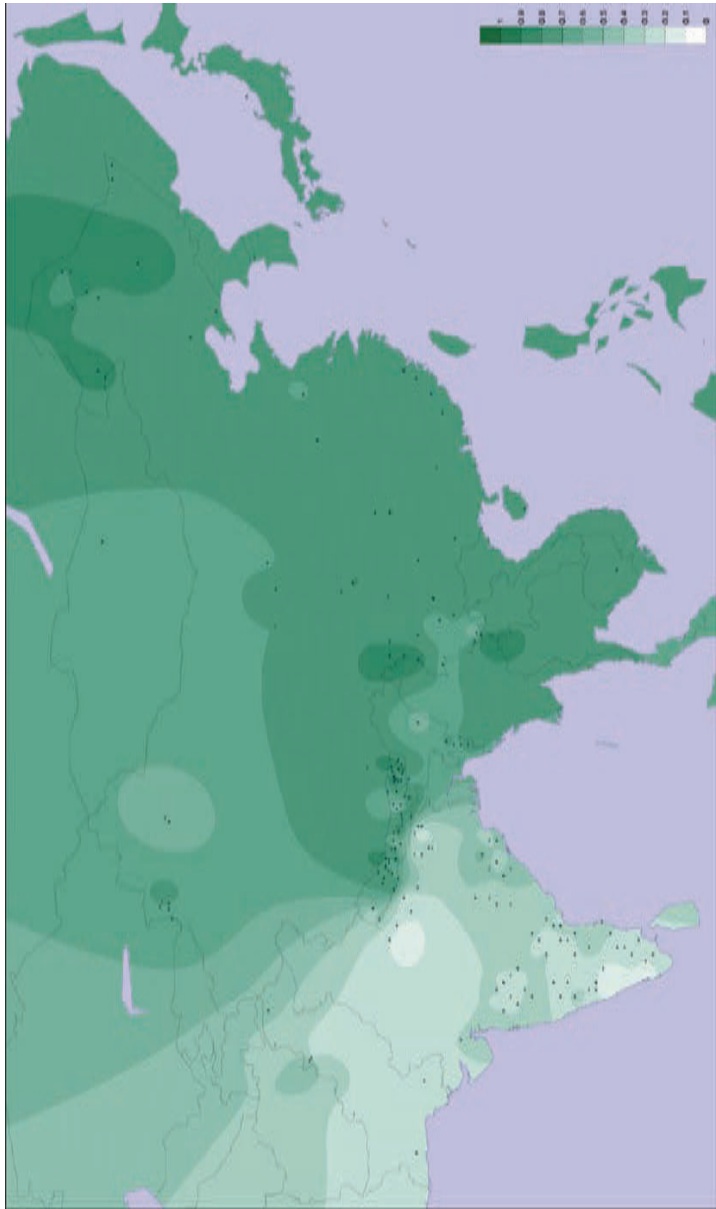


Figure 3. Geographical mapping of the unsupervised Structure K=2 results

The spatial mapping of the Structure K=2 results illustrates an approximate north-east vs. south-west clinal gradient with the steepest gradient located in the southern Himalayan foothills. The green colour scale indicates the percentage of “Tibeto-Burman” genetic contribution which is the highest in Eastern China, and the lowest in Southern India.

The spatial mapping of the Structure $K=2$ results (Figure 3) illustrates an approximate north-east vs. south-west clinal gradient with the steepest gradient located in the southern Himalayan foothills.

In order to compare with the results obtained using Structure and make a more detailed comparison of the sampled populations, pairwise F_{ST} values were calculated and used in multi-dimensional scaling analyses. Figure 4 shows the MDS plot of the first two dimensions, with the populations coded according to language affiliation (see figure legend for explanation of the symbols). As can be seen from Figure 4, there is again a clear subdivision between Tibeto-Burman and Indo-European languages with most of the Nepalese and Bhutanese Tibeto-Burman populations clustering closely with the majority of the Tibeto-Burman reference populations, thus indicating that the genetic distances observed between the populations in this study are correlated more with linguistic distance than with geographic distance.

Even though autosomal STRs are usually thought not to be the best tools for a refined genetic analysis, our study shows that, at least in the Greater Himalayan Region, even a rather small number ($n=15$) of highly-variable autosomal STRs can give a valuable insight into population (pre-) history. Based on initial results from Y-chromosomal and/or mitochondrial analyses (Metspalu et al. 2004, Gayden et al. 2007) it was already suggested that there is evidence for a genetic difference between Tibeto-Burman and Indo-European populations. Our autosomal analyses among a large number of populations from the actual language border area provide good support for this hypothesis. We expect that this will be further confirmed once data from potentially more powerful genetic markers (autosomal SNPs, detailed mtDNA data and detailed Y-chromosome data) become available. These analyses have not been completed yet, but will be available soon.

What are the consequences of the results of the present data for our initial three main research questions? Below we will briefly discuss this:

- Is there a correlation between language, genes, and geography in the Himalayan region?

The answer to this question is a partial yes. We provide evidence, on the basis of autosomal STRs, that there is clear congruence between language and genetics. Populations speaking a language belonging to the Tibeto-Burman language family are genetically more similar to each other than to populations speaking a language belonging to the Indo-European language family. On the basis of language differences we can draw a linguistic boundary roughly running from east to west, just south of the border between India and Bhutan, and running through Nepal. A genetic boundary can be reconstructed along nearly the same route.

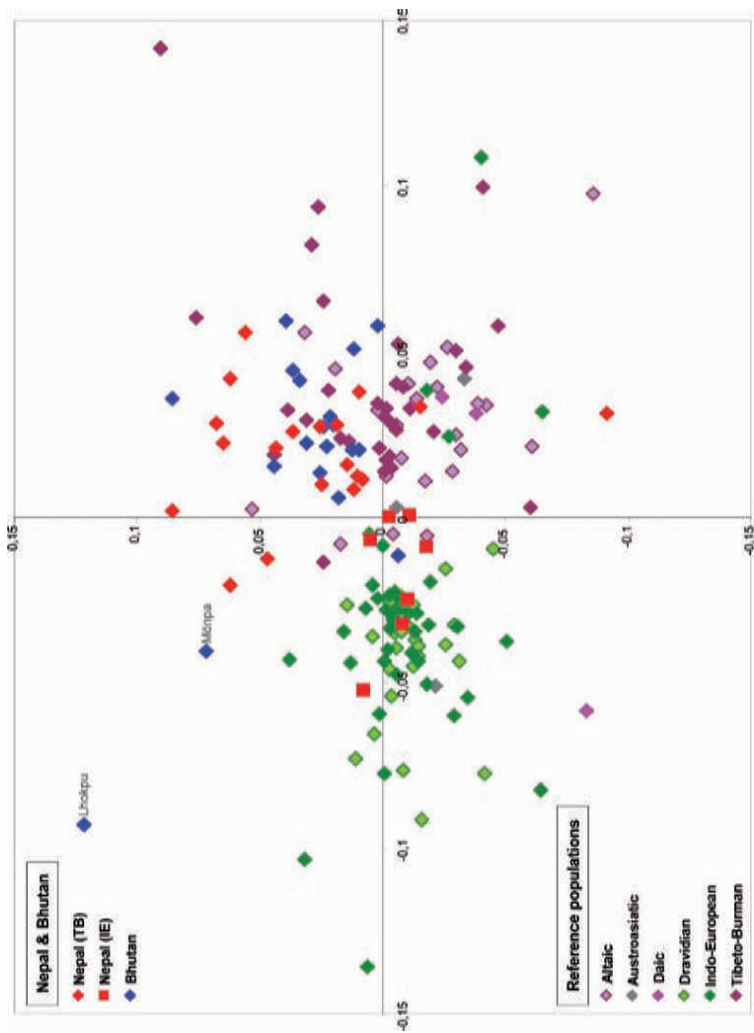


Figure 4. MDS plot

MDS plot, based on pairwise F_{ST} values between all populations. The symbol-coding legend included in the figure reveals the language affiliation of the reference populations and the populations from Nepal and Bhutan. Nearly all Indo-European and Dravidian speaking populations (including those in Nepal) are very clearly separated by the first dimension (x-axis) from populations speaking Tibeto-Burman or other languages.

- Can we determine the genetic relationships (ancient ancestors) of the Nepalese and Bhutanese and deduce possible migration routes?

This question is very difficult to answer on the basis of these autosomal data. Generally speaking we detect a very close genetic relationship between various, mostly eastern Chinese, populations and the Nepalese and Bhutanese populations. However, we can not pinpoint a possible region of origin in China and reconstruct likely migration routes. For this we will need detailed Y-chromosomal data and mtDNA data. In addition, a very large area between the sub-Himalayan region and central China and Tibet has not been sampled. As a consequence, even if we have the Y-data and mtDNA data, we have no information about an essential link between the more eastern Chinese populations and those of Nepal and Bhutan. Because of the marked genetic discontinuity between most of Nepal and Bhutan on one hand and India on the other, we can safely rule out any strong evolutionary (genetic) link between these two countries, except for the Indo-European speaking populations in Nepal. It has been suggested that the extreme northeast of India could have served as a corridor from eastern Asia into India and perhaps Bhutan (Cordeaux et al. 2004). At present we cannot confirm this hypothesis in more detail.

- Can we say something about relative ages of the various groups, identifying and comparing “aboriginal” groups with the others?

In order to answer this question, we would again need more detailed genetic information from neighbouring populations. At a first glance, we do not detect any notable differences between the genetic compositions of Bhutanese language groups and Tibeto-Burman Nepalese language groups. Once we have more information from seemingly isolated aboriginal groups from the north of India, and from isolated language groups from Tibet (the possible direct ancestral source of many of the Nepalese and Bhutanese language groups), we shall be in a much better position to make such inferences.

Acknowledgements

This work, as part of the European Science Foundation EUROCORES Programme OMLL, was supported by funds from NWO (Netherlands Organisation for Scientific Research, grant no. 231-70-001), the Arts and Humanities Research Board of Great Britain (grant no. AN/9585/APN15314) and the EC Sixth Framework Programme (Contract no. ERAS-CT-2003-980409). CTS is supported by The Wellcome Trust. MAJ is a Wellcome Trust Senior Fellow in Basic Biomedical Science (grant number 057559).

The research in Nepal was conducted in association with the Centre for Nepal and Asian Studies (CNAS) at Kirtipur under the Bilateral Agreement for Academic Cooperation between Tribhuvan University (TU) and Leiden University. The research team received the organisational support and enthusiastic assistance of many grass-roots community service organisations and from the many informed volunteers of indigenous language communities who took an active interest in the research and discussed with interest the ramifications of genetic investigations for an enhanced understanding of our shared prehistorical past. Special gratitude is due the Arjun Limbu, Yograj Limbu and the Kirat Yakthung Chumlung headquartered at Mahalakshmisthan in Lalitpur; the Praja Capacity Development Programme (Praja Samudayik Vikash Karyakram) at Shaktikhor in Chitwan district; to Ajay Praja, Santa Bahadur Praja and Dambar Bahadur Chepang and the Nepal Chepang (Praja) Sangh at Kathmandu; to our many Tharu friends at Sauraha; to Bal Gopal Shrestha of Sankhu and the Newar community organisation Friends of Sankhu; to Bharat Rai of the National Youth Service Trust, our dear friend Ashok Rai and the Danuwar Rai community; to Dileन्द्र Subba and the Limbu Literary Development Association; to the Kirat Yakkha Chuma headquartered at Mahalakshmisthan in Lalitpur; to Shree Mani Chand Chantyal and the Nepal Chantyal Samaj headquartered in Samakhushi in Kathmandu; to Buddhiman Dura, Lt. Col. John P. Cross, Ritu Kumar Dura and the Dura Seva Samaj Sampark Karyalaya at Ram Bajar in Pokhara; to Kishor Dura, Singh Raj Dura and the Dura Seva Samaj Samiti at Vasundhara in Kathmandu; to the Kshetriya Karyalaya of the Thakali Seva Samiti at Nadipur in Pokhara; to Dashrath Rai and the Kendriya Karyalaya of the Kirat Rai Yayokkha at Kathmandu; to Lile Thangmi, Kavi Raj Thangmi of Lapilang and the Nepal Thami Samaj; to Avinath Rai and Ganesh Rai of the Wambule Rai Samaj Nepal; to our many kind Baram friends of Gorkha district, for whom we are still completing a grammatical description of the language; to our wonderful friend Temba Bhote and the Buddhist half of the Lhomi (Shingsaba, Bhote) community; to Lt. Col. Michael Roe, Capt. Simon Garside and those troops of the British Gurkhas at Pokhara who came forward to volunteer blood; to our old friend Vishva 'Bishow' Bhatt of Gorkha; to Tek Bahadur Kumal and the Kumal Service Organisation at Chevetar in Gorkha district; to the kind director and staff of the Youth Awareness Environmental Forum and the Environmental Library at Badegaon at Godawari. Gratitude is also due to the Nepal Janjati Mahasangh at Anamnagar for their scholarly interest and kind advice. Furthermore, we owe gratitude to Cas F. de Stoppelaar, Honorary Consul-General of Nepal in the Netherlands and to Kari Cuelenare of the Dutch Consulate in Nepal. Last but not least, we thank our old and dear friends Narayan Prasad 'Yangsarumba' Pangyangu Subba and Gram

Bahadur 'Sarumba' Pangyangu Subba of Tamphula in Tehrathum district and our dear friend Surendra Raj Dhakal of Gorkha.

The research in the Kingdom of Bhutan was conducted as part of the long-standing bilateral cooperation between the Royal Government of Bhutan and Leiden University. The field campaign was carried out in accordance with the Memorandum of Understanding concluded between the Dzongkha Development Authority (DDA) and Leiden University. Much gratitude is due to the Chairmen of the Dzongkha Development Authority, Their Excellencies 'Lönpo Sangay Ngedup and 'Lönpo Thinley Gyamtsho, as well as to the Home Minister, His Excellency 'Lönpo Jigmi Yoezer Thinley, for encouraging and facilitating the research in every way. Logistics and preparations at district and village level were coordinated in a timely and thorough fashion by Dr'asho Sangye Dorji, the Honourable Secretary of Dzongkha Development Authority, and by Tshewang Dorji, the Chief Research Officer.

Much gratitude is due to all the village headmen and district officers who assisted the research team on site. Particular gratitude is due to our friends 'Adap Dôji and Seta of the Lhokpu language community at Loto Kucu, Karma Chen and Dhani Ram Toto at Phüntsho'ling, Tandri and friends at Riti in the Black Mountains, 'Ap Drakpa of Phajong Pam and friends in the Gongduk language community. Furthermore, we thank Cecila Keijzer, Peter Newsum and Yanchen Doma of SNV (Netherlands Development Organisation), Bhutan. During the two expeditions we were assisted in the field by Ivo van Asperen, Janine van Nes, and Dr. Mariëtte Hoffer. Some data, included in this manuscript, and to be published in detail elsewhere (Kraaijenbrink et al. in prep.) contains genetic data provided to us by V.K. Kashyap, B. Su, H. Shi, C.J. Xiao, W.R. Tang, and Y. Xue. Here we would like to acknowledge them for this. Special thanks are due to our old friend Karma Tshering of Gaselô, who really made the work happen. Karma involved himself and made himself indispensable at every level, anticipated and prevented all possible difficulties, and facilitated every aspect of the research programme with unparalleled perspicacity. Finally we would like to thank Mr. H.N.W. van Gent of the Netherlands Ministry of Foreign Affairs in The Hague for his assistance in sending the materials to Nepal and Bhutan.

References

- Cavalli-Sforza, Luigi L., Menozzi, Paolo and Piazza, Alberto. 1994. *The History and Geography of Human Genes*. Princeton, New Jersey: Princeton University Press.
- Cordaux, Richard, Weiss, Gunter, Saha, Nilmani and Stoneking, Mark. 2004. "The northeast Indian passageway: a barrier or corridor for human migrations?" *Molecular Biology and Evolution* 21 (8): 1525–1533.
- van Driem, George L. 2001. *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region, containing an Introduction to the Symbiotic Theory of Language* (2 vols.). Leiden: Brill.
- Falush, Daniel, Stephens, Matthew and Pritchard, Jonathan K. 2003. "Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies." *Genetics* 164: 1567–1587.
- Falush, Daniel, Stephens, Matthew and Pritchard, Jonathan K. 2007. "Inference of population structure using multilocus genotype data: dominant markers and null alleles." *Molecular Ecology Notes* 7: 574–578.
- Gaikwad, Sonali and Kashyap, V.K. 2003. "Genetic diversity in four tribal groups of western India: a survey of polymorphism in 15 STR loci and their application in human identification." *Forensic Science International* 134 (2–3): 225–231.
- Gayden, Tenzin, Cadenas, Alicia M., Regueiro, Maria, Singh, Nanda B., Zhivotovsky, Lev A., Underhill, Peter A., Cavalli-Sforza, Luigi L., and Herrera, Rene J. 2007. "The Himalayas as a directional barrier to gene flow." *American Journal of Human Genetics* 80: 884–894.
- Hima Bindu, G., Trivedi, R. and Kashyap, V.K. 2005. "Genotypic polymorphisms at seventeen autosomal short tandem repeat loci in four tribal populations of Andhra Pradesh, India." *Journal of Forensic Sciences* 50 (4): 978–983.
- Hima Bindu, G., Trivedi, R. and Kashyap, V.K. 2007. "Allele frequency distribution based on 17 STR markers in three major Dravidian linguistic populations of Andhra Pradesh, India." *Forensic Science International* 170 (1): 76–85.
- Kashyap, V.K., Ashma, Richa, Gaikwad, Sonali, Sarkar, B.N. and Trivedi, R. 2004. "Deciphering diversity in populations of various linguistic and ethnic affiliations of different geographical regions in India: analysis based on 15 microsatellite markers." *Journal of Genetics* 83 (1): 49–63.
- Kashyap, V.K., Guha, Saurav, Sitalaximi, T., Hima Bindu, G., Hasnain, Seyed E. and Trivedi, R. 2006. "Genetic structure of Indian populations based on fifteen autosomal microsatellite loci." *BMC Genetics* 7: 28.
- Kraaijenbrink, Thirsa, van Driem, George L., Opgenort, Jean-Robert M.L., Tuladhar, Nirmal M. and de Knijff, Peter. 2007. "Allele frequency distribution for 21 autosomal STR loci in Nepal." *Forensic Science International* 168 (2–3): 227–231.
- Kraaijenbrink, Thirsa, van Driem, George L., Karma Tshering of Gaselô and de Knijff, Peter. 2007. "Allele frequency distribution for 21 autosomal STR loci in Bhutan." *Forensic Science International* 170 (1): 68–72.
- Krithika, S., Trivedi, R., Kashyap, V.K. and Vasulu, T.S. 2006. "Antiquity, geographic contiguity and genetic affinity among Tibeto-Burman populations of India: a microsatellite study." *Annals of Human Biology* 33 (1): 26–42.

- Lee, Andrew C., Kamalam, Angamuthu, Adams, Susan M. and Jobling, Mark A. 2004. "Molecular evidence for absence of Y-linkage of the Hairy Ears trait." *European Journal of Human Genetics* 12: 1077–1079.
- Metspalu, Mait, Kivisild, Toomas, Metspalu, Ene, Parik, Jüri, Hudjashov, Georgi, Kaldma, Katrin, Serk, Piia, Karmin, Monika, Behar, Doron M., Gilbert, M. Thomas P., Endicott, Phillip, Mastana, Sarabjit, Papiha, Surinder S., Skorecki, Karl, Torroni, Antonio and Villems, Richard. 2004. "Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans." *BMC Genetics* 5: 26.
- Neeta, Sarkar and Kashyap, V.K. 2004. "Allelic variation at 15 microsatellite loci in one important Australoid and two Indocaucasoid groups of Chhattisgarh-India." *Journal of Forensic Sciences* 49 (1): 184–188.
- Peakall, Rod and Smouse, Peter E. 2006. "GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research." *Molecular Ecology Notes* 6: 288–295.
- Pritchard, Jonathan K, Stephens, Matthew and Donnelly, Peter. 2000. "Inference of population structure from multilocus genotype data." *Genetics* 155: 945–959.
- Quintana-Murci, Lluís, Krausz, Csilla, Zerjal, Tatiana, Sayar, S. Hamid, Hammer, Michael F., Mehdi, S. Qasim, Ayub, Qasim, Qamar, Raheel, Mohyuddin, Aisha, Radhakrishna, Up-pala, Jobling, Mark A., Tyler-Smith, Chris and McElreavey, Ken. 2001. "Y-Chromosome lineages trace diffusion of people and languages in Southwestern Asia." *American Journal of Human Genetics* 68: 537–542.
- Rajkumar, Ravathi and Kashyap V.K. 2003. "Evaluation of 15 biparental STR loci in human identification and genetic study of the Kannada-speaking groups of India." *American Journal of Forensic Medicine and Pathology* 24 (2): 187–192.
- Xue, Yali, Zerjal, Tatiana, Bao, Weidong, Zhu, Suling, Shu, Qunfang, Xu, Jiujin, Du, Ruofy, Fu, Songbin, Li, Pu, Hurles, Matthew E., Yang, Huanming and Tyler-Smith, Chris. 2006. "Male demography in East Asia: a North-South contrast in human population expansion times." *Genetics* 172: 2431–2439.

Table 1. Descriptive statistics of the population samples from Nepal and Bhutan

Population / Pool	Code	Language family	country
Barâm	BAR	TB	Nepal
Chantyal	CHN	TB	Nepal
Chepang (Praja)	CHP	TB	Nepal
Central Kiranti*	CKI	TB	Nepal
Dhimal	DHI	TB	Nepal
Dura	DUR	TB	Nepal
Eastern Kiranti**	EKI	TB	Nepal
Ghale	GHL	TB	Nepal
Gurung	GUR	TB	Nepal
High Caste Newar	HCN	TB	Nepal
Kham (Magar)	KHM	TB	Nepal
Limbu†	LIM	TB	Nepal
Magar	MGR	TB	Nepal
Newar	NWR	TB	Nepal
Sherpa (Solu-Khumbu)	SHE	TB	Nepal
Thangmi	THG	TB	Nepal
Thakali	THK	TB	Nepal
Tamang	TMG	TB	Nepal
Western Kiranti‡	WKI	TB	Nepal
Bahun (Brahmin)	BHU	IE	Nepal
Chetri (Kshetriya)	CHE	IE	Nepal
Indo European / Tibeto Burman substrate§	IET	IE	Nepal
Indo European¶	IEU	IE	Nepal
Kumal	KUM	IE	Nepal
Majhi (Bote)	MAJ	IE	Nepal
Tharu	THR	IE	Nepal
Brokpa (Bj'op)	BRP	TB	Bhutan
Bumthang	BUM	TB	Bhutan
Chali	CHL	TB	Bhutan
Dakpa (Dwagspo)	DAK	TB	Bhutan
Dzala	DZA	TB	Bhutan
Gongduk	GNG	TB	Bhutan
Brokkat	KAT	TB	Bhutan
Khengpa	KHG	TB	Bhutan
Kurtöp	KUR	TB	Bhutan
Lakha	LAK	TB	Bhutan
Layap	LAY	TB	Bhutan
Lhokpu (Lhop, Doya)	LHP	TB	Bhutan

n#males	n#females	Number on map	Lat (dec)	Long (dec)
32	6	8	28,07	84,67
21	2	2	28,40	83,37
20	7	9	27,58	84,70
42	6	16	27,13	87,05
20	2	19	26,50	87,70
27	8	6	28,28	84,20
12	7	17	27,14	87,43
17	8	7	28,28	84,73
40	6	5	28,30	84,12
24	6	12	27,62	85,43
13	1	1	28,50	83,00
56	7	18	27,19	87,83
40	6	4	28,08	83,83
44	10	11	27,62	85,40
20	5	14	27,73	86,58
16	2	13	27,75	86,00
20	9	3	28,82	83,75
41	9	10	27,88	85,42
51	14	15	27,38	86,60
25	8	21	29,17	81,17
37	10	22	29,17	81,20
33	6	26	27,25	85,75
26	14	20	28,75	80,50
21	5	25	28,05	84,45
21	6	24	27,83	83,67
28	7	23	27,42	83,33
40	10	46	27,40	91,72
50	10	37	27,67	90,55
50	11	42	27,38	91,02
49	10	45	27,47	91,52
51	11	43	27,90	91,15
46	10	41	27,08	90,93
24	5	38	27,73	90,43
52	10	39	27,132	90,68
51	13	40	27,82	90,82
50	10	32	27,68	90,15
25	5	30	28,15	89,40
39	8	29	26,95	89,12

Population / Pool	Code	Language family	country
Mangde (‘Nyenkha, Henke)	MNG	TB	Bhutan
Black Mountain Mönpa	MON	TB	Bhutan
‘Ngalop (Dzongkha)	NGA	TB	Bhutan
Nup	NUP	TB	Bhutan
Tshangla (Sháchop)	TSH	TB	Bhutan
Bodo	BOD	TB	India
Toto	TOT	TB	India
Darai	DAR	IE	Nepal
Giri	GIR	IE	Nepal
Jirel	JIR	TB	Nepal
Lhomi (Shingsaba)	LHM	TB	Nepal
Rana	RAN	IE	Nepal
Shah	SHH	IE	Nepal
Tibetan	TIB	TB	Nepal

* Pool containing Bantawa, Chintang, Chamling, Dungmali, Kulung, Nachiring, Puma and Sampang population samples.

** Pool containing Athpahariya, Lohorung and Yakkha population samples.

† Pool containing Chathare, Pañcthare, Phedappe, Tamarkhole and general Limbu population samples.

‡ Pool containing Bahing, Dumi, Jero, Khaling, Sunwar, Thulung and Wambule population samples.

§ Pool containing Danuwar and Kachariya Danuwar populations samples.

¶ Pool containing Damai, Sarkí, Sonar and Visvakarma population samples.

n#males	n#females	Number on map	Lat (dec)	Long (dec)
54	10	33	27,42	90,22
40	18	34	27,22	90,22
50	10	31	27,53	89,48
27	10	35	27,58	90,33
50	11	44	27,18	91,32
37	2	36	26,67	90,33
54	16	28	26,67	89,00
0	2	small sample size, not included in this study		
3	1	small sample size, not included in this study		
2	0	small sample size, not included in this study		
10	2	small sample size, not included in this study		
1	0	small sample size, not included in this study		
0	1	small sample size, not included in this study		
1	0	small sample size, not included in this study		

Language acquisition and language universals

From babbling to first words in four languages

Common trends across languages and individual differences

Sophie Kern¹, Barbara Davis² and Inge Zink³

The aim of this paper is to compare the developmental trajectory of speech production capacities in children acquiring different languages from the babbling period to the emergence of early grammar. The question of phonetic continuity between babbling and the early lexicon was addressed with the following questions: Do children across languages follow universal trends described previously? Are sound patterns used in babbling the same as in first words across languages? How do child patterns compare with ambient language patterns?

We collected and imposed a uniform analysis profile on large corpora for four diverse languages: French, Romanian, Dutch and Tunisian Arabic. Fifteen children developing normally in a monolingual environment according to community standards and reports from parents and physicians regarding developmental milestones were observed in their typical daily environment. One hour of spontaneous vocalization data was audio and video recorded every two weeks from 8 through 25 months. 529 hours were recorded and transcribed using International Phonetic Alphabet conventions. Lexical and phonetic characteristics were longitudinally analyzed for each of the four languages. Second, minimally 1,000 dictionary entries from the ambient language employed by the parents of each child participant were analyzed for comparison with the child data for that ambient language. Our consistent data collection methods allow consideration of common trends, individual and cross-linguistic language differences in child and in ambient language databases.

Despite individual differences found within and across languages, the majority of children tended to follow the common trends reported in the literature. Our data allowed us also to consider with the question of continuity across babbling, first word, and later word periods. Results indicated that phonetic complexity

-
1. Laboratory Dynamique du Langage, Lyon, France
 2. The University of Texas at Austin, Austin, Texas, USA
 3. Exp. ORL, Dept. Neurosciences, K.U. Leuven, Belgium

expands after the lexical spurt. This increase in complexity could be related to increasing control over speech articulators as well as increase in cognitive and perceptual capacities for matching language forms. To answer our second question regarding the influence of ambient language input, we compared child frequencies to patterns in the 1,000 word dictionary data for each language. These results indicate that input does not play an obvious role in observed patterns across these languages; children are more like one another across languages than they are like their ambient language in this early period.

Introduction

Emergence of complexity in modern child speech output affords a unique opportunity from the perspective of contemporary ontogeny to view potential correlates of complexity emergence across the deep historical process of language phylogeny. Modern language illustrates a pinnacle achievement of system based complexity. This complex system enables humans to achieve mutually intelligible communication using a symbolic system to code an infinite array of topics. Linguistic communication is further shaded by the nuances of linguistic, cultural and social norms in the 5–6,000 modern languages spoken today. Phonological systems in these widely diverse modern languages and language families illustrate the vast array of choices made by speakers throughout the phylogenetic process in implementation of vocal system capacities for sending and receiving messages. In the deep history of language evolution, speakers in diverse communities implemented distinct and disparate capacities of the production system in service of message transmission. These diverse systems of ambient language phonological regularity were driven by the need to increase the complexity of messages sent and received in the vocal medium as social cultures grew in complexity as well (see d'Errico et al., this volume, for a perspective from archaeological investigations).

In modern infants, early speech output patterns based on maturation of immature physical production systems are intertwined with emergence of sound patterns reflecting phonological distinctions for coding messages. Ambient language differentiation is based on learning processes as well as on implementation of more complex patterns produced by the child in the speech output system. Similar to the phylogenetic process, emergence of ambient language speech production patterns in ontogeny shows movement from simple vocal output systems at the onset of speech like canonical babbling (Kern & Davis, 2009) toward more complexity of movement patterning as infants gain cognitive, motor and perceptual capacities to send and receive increasingly complex messages (see Vihman, 1997, for a review). One primary question arises: when do infants begin to display in output the pre-

cise regularities particular to their own ambient language? An additional question in understanding emergence of complexity concerns the order in which components of vocal output related to ambient language regularities emerge. Order of emergence of these precise phonological regularities unique to each infant's ambient language community is founded on multiple variables. Paramount foundations of learning of ambient language regularities in modern infants include the status of speech motor output systems (e.g., Green et al., 2002; Boliek et al., 1997) as well as perceptual capacities for learning nuanced ambient phonological characteristics from adult ambient language input (e. g., Werker & Curtin, 2005).

Of considerable importance as well in understanding potential connections of ontogenetic findings to phylogeny is the relationship of cognitive capacities to emergence of systematic phonological knowledge and behavioural capacities. Individual differences in infant developmental trajectories within and across languages provide a critical dimension of inquiry needed for understanding the course of ontogeny comprehensively. To the extent that children's vocal output patterns are more diverse within and across languages than they are consistent, cognitive processes are implicated in understanding the course of phonological development. In contrast, consistent vocal output trends with little individual variation implicates production system constraints in the context of early emergence of perceptual learning capacities (Saffran, Aslin, & Newport, 1996).

Available studies of early periods of ontogeny relative to common vocal output trends and emergence of ambient language learning are prominently centered on the canonical babbling and early single word periods of development. These studies indicate common effects in production patterns as well as some indication of early appearance of output patterns related to input from the ambient language. To understand the interaction of patterns based on production system propensities common to all infants versus the role of early learning from unique environmental regularities in the ambient phonology, it is important to examine diverse languages that produce targets for learning outside the common characteristics suggested as being present in many infants across languages. That is a major goal of this program of study. The larger philosophical goal is to consider the nature of these patterns as they illuminate understanding of the process and the behavioral correlates of emergent complexity in the evolution of the vocal capacity in early speakers.

Early development of speech production capacities in modern infants

Prelinguistic vocalizations of modern infants, in the period before language emerges have been frequently studied. These earliest speech like patterns, termed canonical babbling (Oller, 2000), are perceptually similar to early language based speech output but do not code meanings. These vocalizations show common pat-

turning across languages (Kern & Davis, 2009). A near universal basis for babbling has been postulated based on available studies. Similarities in sound types (i.e. “b” or “d”), sound combinations (e. g., “ba”, “da”, “ku”) and utterance types (e. g. simple CV or CVCV utterances such as “ba” or “dadi”) across different language communities have been frequently documented.

Consonants most frequently reported are stops (total occlusions, e. g., “p”), and to a lesser extent, nasals (e. g., “m”) and glides (e. g., “w”) (Locke, 1983; Roug, Landberg & Lundberg, 1989; Stoel-Gammon, 1985; Vihman *et al.*, 1985). During the babbling period, children tend also to produce many coronals (tongue tip, e.g., “d”) and labial (lip closure, e. g., “b”) consonants (Locke, 1983) and few dorsals (tongue back, e. g., “k”; Stoel-Gammon, 1985). Vowels from the lower left quadrant of the vowel space (i.e. mid and low front and central vowels; e. g., “ah” or “uh”) are most often observed (Bickley, 1983; Buhr, 1980; Kent & Bauer, 1985; Lieberman, 1980; MacNeilage & Davis, 1990; Stoel-Gammon & Harrington, 1990). Languages exhibit great diversity in consonant and vowel sound types, but coronal place and stop and nasal manner of articulation are present and predominant in a large number of world languages (Ladefoged & Maddieson, 1996). Oral occlusives are more frequent than nasals in child repertoires and in language inventories. Front and central vowels are also prominent in inventories across languages, although vowel diversity is quite broad in languages compared with early child production patterns.

Children tend to favor open (CV, e.g., “da”) as opposed to closed (CVC, e.g., “cot”) syllable types (Kent & Bauer, 1985; Locke, 1983; Oller & Eilers, 1982; Stoel-Gammon, 1985; Vihman, 1992). These common infant preferences, found in babbling and first words, are retained in some languages. The open syllable is the only universal syllable type in languages (Maddieson, 1984).

Speech sounds in vocal utterances are not produced alone but as a series of alternations of consonants and vowels. This phenomenon of serial ordering is one of the most distinctive properties of speech production. MacNeilage and Davis (1993, 2000) have tested a model termed the “Frame-Content Theory” to characterize the serial organization of vocalizations in babbling (Davis & MacNeilage, 1995), first words (Davis, MacNeilage, & Matyear 2002), and languages (MacNeilage, Davis, Kinney, & Matyear, 2000).

In the Frame-Content Theory, biomechanical based predictions have been made regarding both intrasyllabic and intersyllabic regularities in vocal output. Intrasyllabic consonant-vowel alternations form the basis for syllables in spoken language. In pre-linguistic canonical babbling, they are proposed as being based on rhythmic close-open mandibular oscillations accompanied by phonation. These rhythmic mandibular oscillations (termed “frames”) comprise the production system based unit within which speech-like behaviors first emerge. As children gain control over the independent coordination of articulators (including tongue, velum,

and lips) within utterances, individual speech sound movements (i. e., “content”) are gradually differentiated to become separate entities in production and perception.

In babbling and first words, “frame dominance”, without independent control of other active articulators, produces predictions for patterns within spoken utterances. Three preferred intrasyllabic co-occurrence patterns emerge from this hypothesis. These co-occurrences are for coronal (tongue tip closure) consonants to occur with front vowels (e.g. “di”), dorsal (tongue back closure) consonants to occur with back vowels (e. g. “ku”), and labial (lip closure) consonants to occur with central vowels (e.g. “ba”). In each of these intrasyllabic sequences, the tongue does not move independently of the jaw within the syllable, but stays in the same position for the open portion (vowel) of the syllable as the jaw opens as it was for the consonant closure portion.

Recent studies have tested predicted serial patterns within utterances in the babbling and early speech period in English and a variety of other languages. In an English language environment studies of 6 children during babbling (Davis & MacNeilage, 1995) and 10 children in first words (Davis, MacNeilage & Matyear, 2002), confirmed predicted patterns at above chance levels using chi square analysis. Predicted intersyllabic and intrasyllabic trends have also been found in analyses of 5 French children, 5 Swedish children and 5 Japanese children from the Stanford Child Language Database (Davis & MacNeilage, 2000), one Brazilian-Portuguese learning child (Teixiera & Davis, 2002), 7 children acquiring Quichua (Gildersleeve-Neumann & Davis, 1998) and 7 Korean children (Lee, Davis & MacNeilage, 2007). While these patterns have not been confirmed universally in these studies, they are present in every language and account for the predominant trends in serial output organization when the corpora are large enough to test predictions.

The very strong persistence of some child patterns in languages as well as their presence at the onset of speech-like vocalizations in infants indicates that they reflect fundamental properties of the speech production system. When patterns are retained in languages, they do not merely represent an aspect of the acquisition process disappearing with maturity of the vocal apparatus. Both the common patterns and the language specific variants observed should have significant implications for understanding of the phylogeny of speech production complexity supporting modern language message transmission. It is possible, for example, that patterns common to all or most languages were present in the first spoken language, as they represent fundamental properties of operation of the production system. In contrast, language specific patterns are perhaps more likely to have emerged later in the phylogenetic process as individual languages diversified to increase the size of their respective message sets.

Perceptual influences on early speech production preferences

In addition to the strong presence of near universal trends at the onset of speech like production output, input from the ambient language plays a role in children's very early perception as early as 8–10 months (e. g., Werker & Tees, 2005). Input from the ambient language has been proposed as shaping children's production preferences at some point in the late babbling and first word periods. This potential for ambient language influence has been examined for utterance and syllable structures (de Boysson-Bardies, 1993; Kopkalli-Yavuz & Topbas, 2000), vowel and consonant repertoires (de Boysson-Bardies, Hallé, Sagart & Durand, 1989; 1984; Lee, Davis & MacNeilage, 2008) and distribution as well as CV co-occurrence preferences (e.g. Lee, Davis & MacNeilage, 2007).

In one representative study, de Boysson-Bardies, Sagart & Durand (1984) presented naïve adults with sequences of early babbling of French, Arabic & Cantonese children. Participants were asked to identify the babbling of French infants. Listeners were correct in judging 70% of tokens, suggesting that babbling in the pre-linguistic period may show language specific intonation characteristics. De Boysson-Bardies, Hallé, Sagart & Durand (1989) also compared vocal sounds of French, English, Cantonese and Algerian infants. They proposed that the acoustic vowel distribution was significantly different for the 4 language groups. De Boysson-Bardies, Hallé, Sagart & Durand (1984) showed the same tendency within the infant's consonant repertoires. Their consonant data indicated significant differences in the distribution of place and manner of consonant articulation across the languages analyzed. French infants produced more labials than Japanese or Swedish infants. De Boysson-Bardies (1993) also examined syllable types. She compared Yoruba babbling to French, English, and Swedish babbling. She interpreted her findings as showing that French, English and Swedish children produce between 65% and 75% of CVCV disyllables, whereas Yoruba children produced 62% VCV word types. She attributed this difference to the particular structure of Yoruba, where most words begin with a vowel.

Levitt & Utman (1992) compared reduplication and variegation in babbling in French and American English-learning infants. They found that French infants showed significantly more reduplicated utterances during the babbling period (four or more syllables in length) than their American-English learning infant cohort. This type of result suggests an early effect of the ambient language related to the length of the prosodic word during the pre-linguistic babbling period.

Present research into early periods of speech ontogeny has most often included studies of small numbers of infants. Diverse methodologies have been employed for analysis. To ensure adequately large corpora of data for statistical analysis of patterns as well as to ensure consistent analysis procedures, we have collected and

imposed a uniform analysis profile on large corpora for five diverse languages. Our goal is to understand common trends as well as the timing and characteristics of appearance of language specific influences in the ontogenetic process of phonological acquisition in modern infants. An additional goal is to observe the process past the earliest periods to begin to understand the emergence of ambient language phonological regularities across diverse languages more fully.

Method

Participants

Fifteen children (nine boys and six girls) selected from four different ambient language environments participated. French (three boys and one girl), Romanian (two boys and one girl), Dutch (two boys and two girls) and Tunisian Arabic (two boys and two girls) participants were included. The languages studied belong to typologically diverse language families: French and Romanian are Romance languages, Dutch is a Germanic language and Tunisian belongs to the Arabic language family.

All children were members of middle to upper class families who were providing monolingual input to their children. Child participants were selected through informal search in each language community. Mothers were expected to participate in the interaction and elicitation of spontaneous language with their children every two weeks across a period of 17 months. Children with siblings as well as only children participated in the study. All participants were reported as being normally developing according to medical and educational reports from parents and caregivers based on community standards. Children were enrolled in the project by the age of eight months when their parents reported that canonical babbling was established. They were followed until the age of 25 months.

Data collection

One focus of the study was analysis of continuity between the phonetic characteristics of speech sounds and sound sequences in babbling with changes across the period of development of the early lexicon to 25 months of age. Data collection started at eight months, the period in which all the children were reported to have started producing canonical babbling. It was terminated at 25 months when all participants (except two Romanian children) were producing a large number of diverse lexical items. Data collected across the entire period was divided into three parts: the *babbling* period between eight months and the acquisition of the first words (around 12 months), the *early word* period between the first word and the

point where the child had acquired a language production of 50 words, and the *later word* period after the acquisition of the first 50 words.

During this data collection time frame of 17 months, one hour of spontaneous language data was recorded every two weeks in the normal home environment of the children. Each infant's vocal output was audio and video recorded.

Mothers participated in interaction during the recordings without any external instruction except to interact with their children as they normally would in daily care routines and in play. In total, 128 hours of babbling were collected; at least six hours per child. For the early word period, 269 hours of recordings were collected, with at least 11 hours per child. In the later word period, 132 hours of data were collected. Two Romanian children did not reach the 50 word level by the age of 25 months. No later word data is available for those two children. Table 1 shows the division of data collection relative to number of one hour sessions.

Table 1 Number of Sessions during Babbling, Early Word and Later Word Periods

Child	Number of sessions by period		
	Babbling	Early words	Later words
French 1	8	19	14
French 2	7	14	10
French 3	9	23	4
French 4	8	19	14
French Total	32	75	42
Romanian 1	8	16	–
Romanian 2	8	21	6
Romanian 3	10	23	–
Romanian Total	26	60	6
Dutch 1	10	18	5
Dutch 2	8	11	17
Dutch 3	10	12	16
Dutch 4	10	15	19
Dutch Total	38	56	57
Tunisian 1	6	15	6
Tunisian 2	10	19	7
Tunisian 3	7	22	8
Tunisian 4	9	22	6
Tunisian Total	32	78	27
Total	128	269	132

Language dictionaries

To compare the spontaneous vocalizations and word based data of the children with the ambient language of their environment, 1,000 entries from the language communities employed by the parents of each infant were collected from dictionaries available on computer. Dictionaries for French, Dutch, Romanian, and Tunisian were employed.

Data transcription

After collection, the child data was phonetically transcribed using the International Phonetic Alphabet (IPA). At the onset of the study, transcribers were trained in infant transcription. Broad phonetic transcriptions were used, supplemented by some diacritics (mainly for palatalized, pharyngealized, nasalized sounds and duration of sounds). Tokens considered as single utterance strings were bounded by one second of silence, noise or adult speech.

Transcribed sounds were perceptually divided into consonants and vowels by transcribers. Consonants were grouped according to 1) place of articulation: *labial* (bilabial, labiodental, labio-palatal and labio-velar), *coronal* (dental, alveolar, post-alveolar and palatal), *dorsal* (velar and uvular) and *guttural* (pharyngeal and glottal) and 2) manner of articulation as oral stops, fricatives, liquids, nasals, glides and “other” (i.e. trills, taps and affricates). Glides were considered as consonants as they share the consonantal property of accompanying the mouth-closing phase of babbling. Vowels were grouped according to 1) height: *high* (high and lowered-high), *mid* (higher-mid, mid and lower-mid) and *low* (raised-low and low) 2) *front/back*: front, central and back. An “other” category included all non-transcribable segments that were not perceptually accessible to transcribers.

All transcribed data were entered for computer analysis using software designed for analysis of infant speech data, particularly the Logical International Phonetic Programs (LIPP, Oller and Delgado, 1990). LIPP is a database program designed to count data that has been phonetically transcribed using IPA symbols.

Data analysis

Six domains were examined for each participant: 1. Frequencies of consonant and vowel sounds, 2. Consonant manner of articulation, 3. Consonant place of articulation, 4. Vowel types, 5. Manner versus place change for consonants in CVCV sequences, 6. Height versus front-back change for vowels in CVCV sequences. To evaluate these structural sound pattern changes across periods several types of analysis were completed.

Means of early appearing sounds or structures were compared to the means of later appearing ones across the three periods in each of the four languages. T-tests were used to assess whether the means of the two groups of sounds or structures were statistically different across the three periods. Alpha level was set at .05. As number of participants in each language was small, with three to four participants, a qualitative analysis of individual child scores was included to understand individual differences within language groups.

A second analysis focussed on cross-language differences. Mean, standard deviation and interval of confidence were calculated for early appearing sounds or structures in each language. In the case of a standard deviation higher than 15 in a language group, the child furthest from the mean was excluded from analysis in order to work with a more homogeneous group. Interval of confidence allows the building of a confidence space for each language. The confidence spaces were then compared in a pair wise comparison. In the case of an overlap between two confidence spaces, the languages concerned were considered as exhibiting the same behaviour. In the case of no overlap, the languages were considered as presenting statistically significant differences.

The goal of a third analysis was to observe changes across the three developmental periods of *babbling*, *early words*, and *later words*. T-tests were used to verify whether the changes observed in sound patterns analyzed from 1) the *babbling* to the *early word* period and 2) the *early word* period to the *later word* period were significantly different from one another in each of the four languages. As it was the case for the first type of analysis, alpha level was set at .05 and a more qualitative analysis of individual children's scores was also conducted.

Results

Consonant and vowel frequencies

Consonants and vowel types transcribed for all participants across the four language groups are enumerated in Table. 2. Table 2 shows that in the *babbling* period, children produced more vowels (60,378) than consonants (51,269). A Wilcoxon signed ranks test indicates that this difference is significant ($p=.009$). In the *early word* period, children produced an approximately equivalent frequency of vowels (21,952) and consonants (20,726). A Wilcoxon signed ranks test shows no significant difference ($p=.427$) between the two sound types. In the *later word* period, children produced more consonants (79,058) than vowels (72,646). Although they produced more consonants, a Wilcoxon signed ranks test showed that the difference was not significant ($p=.279$).

Table 2. Consonant and Vowel Frequency of Occurrence by Period

Consonant and Vowel Frequencies by Period						
	Babbling		Early words		Later words	
Child	Cons	Vowels	Cons	Vowels	Cons	Vowels
French 1	3276	684	1043	1257	7078	7870
French 2	2003	2176	1149	1274	5477	5910
French 3	1464	4073	1934	3005	1725	1912
French 4	3213	4073	948	1131	10026	9545
French Total	9956	11006	5074	6667	24306	25,237
Romanian 1	1558	2026	2813	2265	–	–
Romanian 2	3115	4259	297	260	3804	4001
Romanian 3	2999	3654	1176	1192	–	–
Romanian Total	7672	9939	4286	3717	3804	4001
Dutch 1	5726	7284	2251	2054	1931	1902
Dutch 2	2870	2972	768	565	5442	4018
Dutch 3	1704	2418	1024	1077	10,636	9019
Dutch 4	4419	4885	635	554	17,517	13987
Dutch Total	14,719	17,557	4678	4250	35,526	28,926
Tunisian 1	2625	2995	1254	1028	1197	1033
Tunisian 2	5660	6597	1148	1367	4286	4252
Tunisian 3	4321	4804	2071	2421	8205	7217
Tunisian 4	6316	7480	2215	2502	1734	1798
Tunisian Total	18,922	21,876	6688	7318	15,422	14,300
Total	51,269	60,378	20,726	21,952	79,058	72,464

Consonant manner of articulation

The frequency of stops (oral and nasal) and glides were compared to fricatives, liquids and others (affricates, trill and taps) in the four languages individually, across the four languages and across the three time periods. The consonants have been grouped as function of their order of appearance: the stops and glides being in most cases described as “early appearing” sounds in children’s productions whereas fricatives, liquids and other types of consonants are considered as “later appearing” sounds. Figure 1 presents mean percentage of stops and glides and confidence intervals for French, Romanian, Dutch and Tunisian languages in *babbling* (8 to 12;15 months), *early word* (from the first words to the 50 word-mark) and *later word* (from the 50 word-mark to 25 months) periods. Finally, the mean percentages of

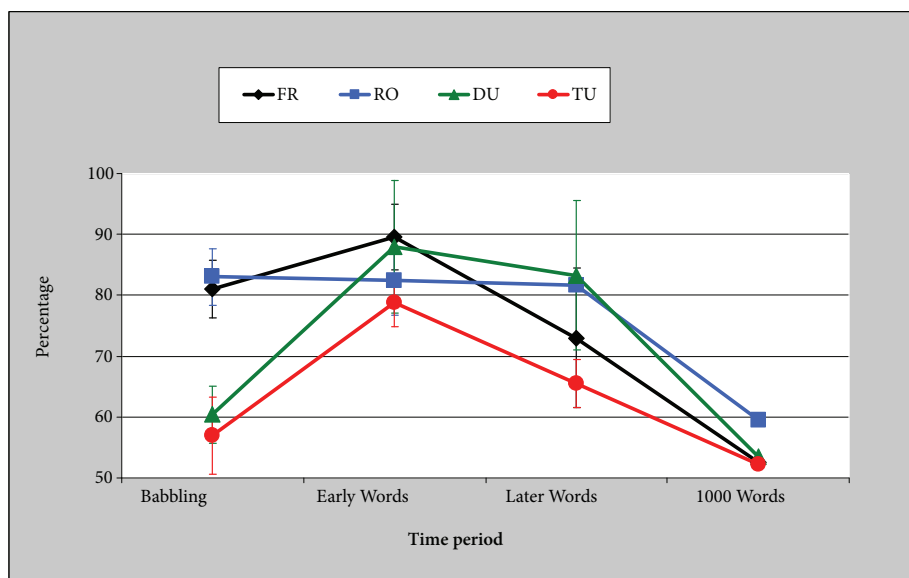


Figure 1. Mean percentage of stops and glides (vs. fricatives, liquids and others) according to language group and developmental period

both categories are given for the 1,000 words which represent the ambient language based on analyses of dictionaries on computer for each language studied.

Participants followed the preferred trends described in previous studies: in all four languages percentage of stops and glides was higher than 50%, indicating a strong preference for early appearing sounds. The differences in frequencies between early and later appearing sounds were statistically significant in all four languages across all three periods. This finding was characteristic of all children except one Tunisian child who used as many early as later appearing sounds in *babbling* and one Romanian child who showed the same pattern during the *early word* period.

Despite these strong general trends, some cross-language differences emerged. During the *babbling* period, the languages were grouped into pairs: Romanian children were like French children in producing more than 80% of stops and glides. Dutch and Tunisian children produced only around 60% of these two manners of articulation. One explanation for these differences might be sought in ambient language input differences. If there were differences in input, these results would indicate early ambient language learning. However, this explanation is not valid relative to Dutch and Tunisian language frequencies. Dutch and Tunisian language analyses did not show lower frequencies of stops and glides than Romanian or French in either sound type (i.e., what sounds occurred) nor in tokens

(i.e., how many times the sounds occurred). Moreover, in the French and Tunisian 1,000 word dictionary analysis, the frequency of stops and glides was not different from the frequency of other manners of articulation whereas in Romanian and Dutch stops and glides were more frequent. The only difference between the two pairs of languages where the children showed diverse patterns is in the number of glottal fricatives. Glottal fricatives are much more frequent in the Dutch and Tunisian languages than in the French or Romanian languages. In the *early word* period, one difference remained: French children used statistically more stops and glides than Tunisian children (89.6% vs. 78.8%). Finally at the *later word* period, Tunisian participants used statistically fewer stops and glides than Dutch children (65.5% vs. 83.3%).

Figure 1 allows also observation of the developmental trajectory from the *babbling* to the *later word* periods. Frequency of stops and glides increased from *babbling* to *early word* periods in three out of four languages and then decreased from the *early word* to the *later word* period. In Romanian, stop and glide manner of articulation decreased significantly from *babbling* to *early words*. As only two Romanian children of three reached the 50 word-mark level, nothing more can be concluded. Most of the children followed this developmental trajectory with an increase of stops and glides followed by a decrease. Only one Romanian child used less stops and glides in the *early word* period than in the *babbling* period, three children used more stops and glides in the *later word* than in the *early word* period (one French, one Romanian and one Dutch) and two produced the same manner patterns in the early and later word periods (one Dutch and one Tunisian).

Consonant place of articulation

Consonant sounds were grouped by place of articulation and by frequency in children's productions: consequently, labials were grouped with coronals as early sound types and dorsals with gutturals as later sound types. Figure 2 presents mean percentage of labials and coronals and confidence intervals for the four languages and the three time periods.

In all four languages, a large majority of children preferred to use labial and coronal consonants over dorsal and guttural consonants across the entire period of observation. All children exhibited this trend during the *early word* and the *later word* periods. Only during *babbling* were some inter-individual differences observed. Two Tunisian children used more dorsal and guttural consonants and two other children (one Dutch and one Tunisian) used as many labials and coronals as other places of articulation.

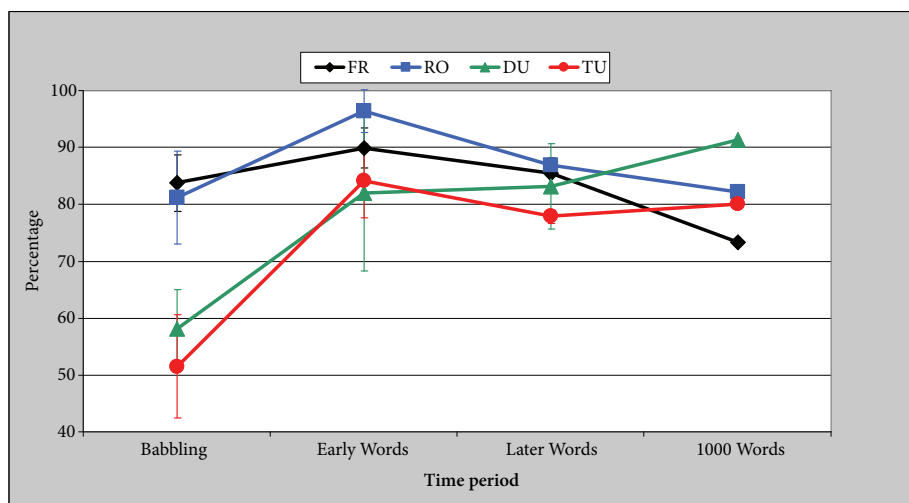


Figure 2. Mean percentage of labials and coronals (vs. dorsals and gutturals) according to language and time period

For cross language comparison, only 14 children were analyzed during the *babbling* period. The standard deviation in Dutch for mean of labials and coronals was higher than 15 exhibiting an important heterogeneity in Dutch children's productions. Consequently the child who was the furthest from the mean was excluded from analysis. As for manner of articulation, the languages were grouped into pairs: French and Romanian (83.7% and 81.2%) produced statistically more labials and coronals than Dutch and Tunisian (58.1% and 51.5%). This outcome could be related to the previous results for manner of articulation. There are more glottal fricatives in the Dutch and Tunisian languages and these fricatives are pronounced in the back of the vocal tract. During the *early word* period, no cross language differences were observed whereas during the *later word* period only French children (85.5%) continued to be different from Tunisian children (77.9%) in use of place of articulation properties.

Consideration of changes from one period to another one also showed some differences. First, frequency of labials and coronals increased significantly in all four languages from the *babbling* to *first word* periods. Only two children behaved differently. One Dutch child produced equivalent frequencies of labials and coronals in both periods. One French child used fewer labials and coronals in the *early word* than in the *babbling* period. Second, frequency of labials and coronals decreased in three out of four languages from *first word* period to *later word* period. For the Romanian children, not enough data were available to evaluate the *later word* period; only one out of three children had data for that period. The decrease

of labials and coronals in comparison to glottals and gutturals was a preferred trend for 10/13 children. Two Dutch children choose the reverse trend and in one Tunisian child, there was no significant change.

Vowel types

Vowels were grouped as belonging to the lower left part of the vocalic space (LLQ) compared to the “other” types of vowels. LLQ vowels are described as being most frequent in early production inventories across a number of studies of this period. For LLQ, the mid and low front vowels as well as the mid and low central vowels were included. Figure 3 shows the results.

In babbling, children in all four languages produced significantly more vowels from the LLQ of the vocalic space. All children except one followed this trend. In *babbling*, one Romanian child was not considered, as the standard deviation for the Romanian group was over 15. This comparison revealed one statistically significant difference: Tunisian children produced more vowels from the LLQ than the other three languages (90.25%). This result cannot be explained by the influence of ambient language: Tunisian is the only language in which the 1,000 word dictionary data are composed of more “other” vowels than vowels from the left inferior part of vocalic space (52% (1,113) vs. 48% (1,021)).

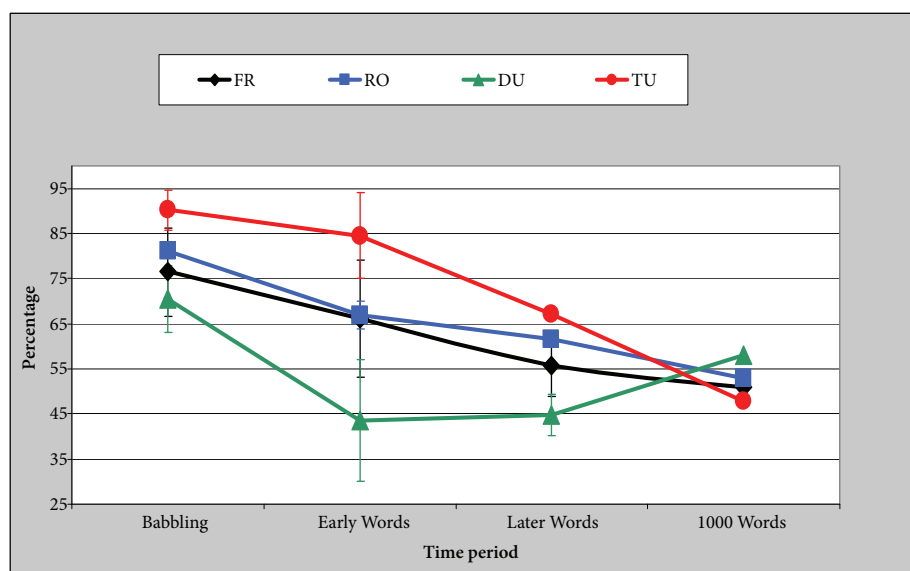


Figure 3. Mean percentage of vowels belonging to the left inferior part of vocalic space (LLQ) vs. “other” vowels according to language and time period

In the *early word* period, French, Romanian and Tunisian children preferred to produce vowels from the LLQ, consistent with *babbling* patterns. Nine of ten children adhered to this preference. In Dutch, only one child showed the LLQ preference, two preferred “other” vowel types and one produced an equal frequency of LLQ vowels and “other” vowels. In cross-language comparisons in the *early word* period, three children (one French, one Romanian and one Tunisian) were not taken into account for analysis as the standard deviation was over 15 in the three language groups. These comparisons revealed a more frequent use of early appearing sounds in Tunisian children as opposed to Romanian and Dutch children (84.5% vs. 67% and 43.5%)

Finally at the *later word* period, as during *early word* period, Dutch children were the only participants not to exhibit a preference for LLQ vowels (45%) as compared to “other” later occurring vowel types. The ambient language learning explanation is not cogent, as the Dutch 1,000 word dictionary analysis showed significantly more early appearing LLQ sounds (1,553 out of 2,690). Three of four French children, the one Romanian child, and all four Tunisian children produced significantly more LLQ early appearing vowels. In contrast, three of four Dutch children preferred “other” later appearing vowels and one produced equal frequencies of the two categories.

A significant decrease in percentages for early appearing LLQ vowels was shown in all four languages across the three periods. Twelve of 15 children showed a decrease in LLQ early occurring vowel types. Only three out of 15 increased their use of LLQ early appearing sounds (one Romanian, one Dutch and one French child). The *later word* period analysis revealed a more heterogeneous use of the vowel space. Dutch children changed to production of more LLQ early appearing vowels during the *later word* period than the *early word* period. In contrast, Tunisian and French children showed continuing decreases in frequency of LLQ early appearing vowels. Nothing could definitively be concluded about Romanian: only one child’s data was available. However, despite these cross language differences, 9/13 children exhibited a decrease in LLQ early appearing vowels from *early word* to *late word* periods. Only two children exhibited an increase in LLQ vowel use and two children exhibited equivalent use of LLQ and “other” vowel types across the three periods analyzed.

Height vs. Front-back changes in CVCV sequences

Variegated babbling in CVCV sequences was analyzed. For vowels, changes in vowel height were compared to changes in the front-back dimension in two successive syllables. According to a large number of studies, children prefer to change the open close dimension of their mouth (producing a vowel height change) than move their tongue into the front-back dimension (producing a vowel front-back

change) across utterances. This prediction emerges from the Frame/Content perspective (MacNeilage & Davis, 1990), where jaw movements are proposed as accounting for across utterance vowel quality changes more than tongue movements independent of the jaw in early utterances. The results are presented in Figure 4.

During the *babbling* period, in all languages except Dutch, children preferred vowel height changes in successive CV syllables significantly more frequently than front/back changes as predicted by the Frame-Content Theory. In Dutch vowel height changes were predominant but the difference with front/back changes was not significant. However, if evaluated individually, only 5/15 children produced significantly more height than front/back vowel changes. The other children followed the same trend but the difference between height and front/back changes was not statistically significant. Cross language comparisons revealed the following differences: Tunisian children produced more vowel height changes than Romanian (73.9% vs. 63.9%). Romanian children then produced more than French (53.4%) and Dutch children (55%). This result could be considered in the context of the highest ratio of height changes to back/front changes in the Tunisian 1,000 word dictionary analysis compared to the other languages (62% height changes vs. 38% back/front changes). However, as this trend is also the most preferred trend in children across these four languages, nothing definitive can be concluded about a possible influence of ambient language.

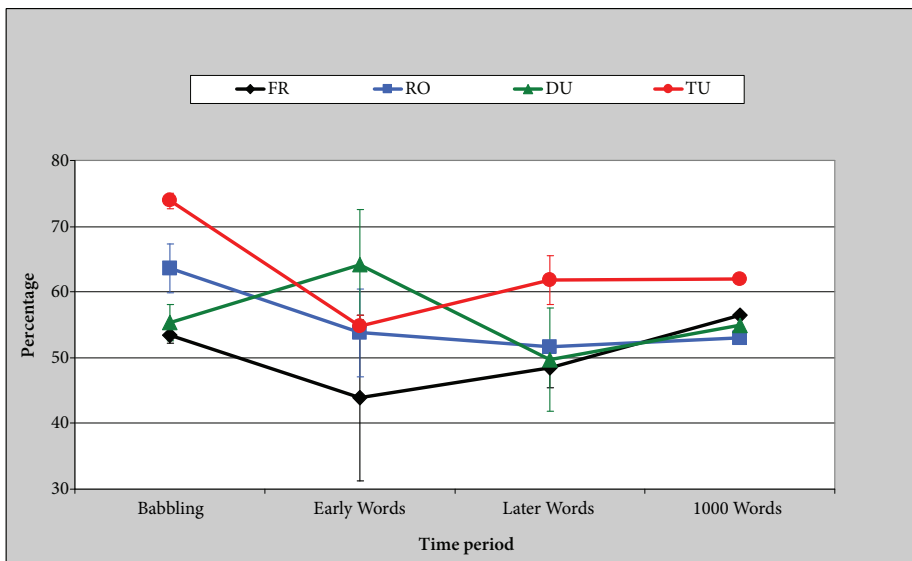


Figure 4. Mean percentage of height changes (vs. front-back changes) according to language and time period

In the *early word* period there was an equal distribution of types of vowel changes in all languages except French. In French, the frequency of front/back changes was statistically higher than height changes (56% vs. 44%). However 10/11 children produced more height changes even if the difference was not statistically significant. One Dutch child had no data of this type in the *early word* period. Only one French child produced more vowel front/back than height changes with a statistically significant difference and one Dutch child exhibited the reverse trend. Contrary to *babbling*, no difference between languages was significant in the *early word* period.

In the *later word* period, French children did not exhibit any preference whereas, in Dutch and Tunisian, vowel height changes were predominant. The only Dutch child with data in this period produced more height changes. 6/13 children (two children had no data) did not exhibit a preference, whereas, 5/13 preferred vowel height changes and only two preferred vowel front-back changes. The only cross language difference to emerge concerned the Tunisian children which, as was already the case in *babbling*, produced more height changes than all other three languages (61.8%).

Longitudinally, from the *babbling* to *early word* periods, a significant decrease of height changes was observed in French and Tunisian and in the one Romanian child (no data were available for two out of three Romanian children). A large majority of children (8/9) revealed the same significant decrease. In Dutch the frequency of height changes increased from *babbling* to *early words* but this was true only in two children out of four (one had no data for early words period and one stayed the same). From the *early word* to the *later word* periods, French and Tunisian children continued to exhibit the same trajectory: a significant increase in height changes. In Dutch, the opposite trajectory was observed. Nothing can be concluded for Romanian as the only child with data for the three periods remained the same. Individual child profiles showed that 6/12 stayed the same, two produced less height changes and 4/12 produced more height changes.

Manner vs. place changes in CVCV sequences

Consonant manner versus place variegation was analyzed in CVCV sequences. Previous studies have shown a preference for manner changes in young children's productions. This is the type of across utterance change predicted by the Frame-Content Theory based on jaw movement without independent movements of other articulators (MacNeilage & Davis, 1990). Consequently, manner changes were considered as the early appearing structures in comparison with place changes. The results are presented in Figure 5.

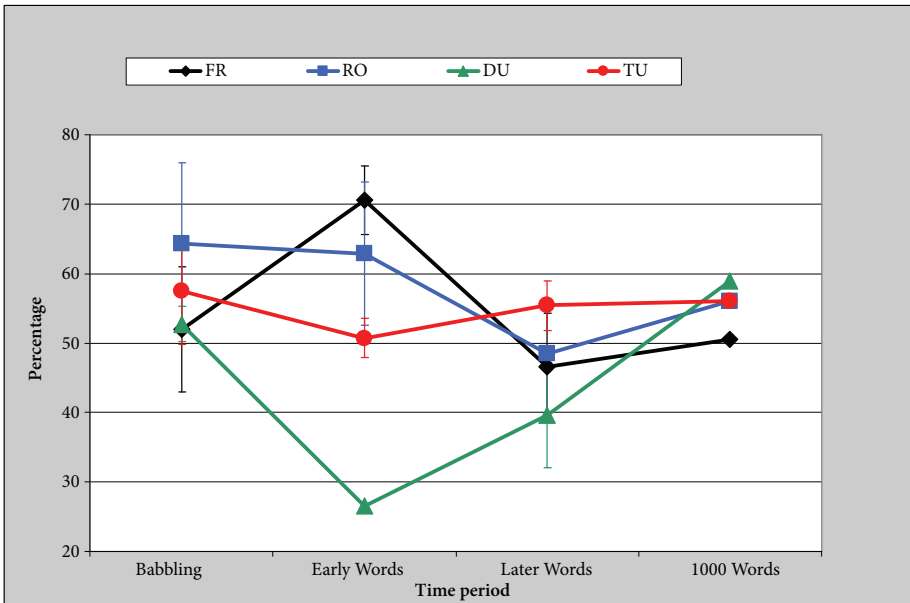


Figure 5. Mean percentage of manner changes (vs. place changes) according to language and time period

As predicted, Romanian and Tunisian children produced significantly more manner than place changes during *babbling* (64.3% and 57.3%). French and Dutch children didn't exhibit any difference between categories. Eleven of 15 children produced as many manner as place changes. Moreover, 10 of the 11 children preferred manner changes even if the difference was not statistically significant. For four children the difference was significant: three used more manner than place changes and one showed the opposite pattern. No significant cross language difference emerged.

In the *early word* period, Romanian, Tunisian and French children produced as many manner as place changes. Eight of 10 children followed this trend. One French child produced significantly more manner than place changes. The only Dutch children with these types of data in the *early word* period produced more place than manner changes. In the cross language analysis, three of four French and two out of three Romanian children were analyzed because of heterogeneity in both groups. The comparison between languages showed a difference between the French and the Tunisian groups: Tunisian children produced more manner changes than the French children. This difference cannot be understood by input frequency. The ambient languages do not present any differences between manner and place changes.

In the *later word* period, differences between groups were particularly important. French and Dutch used more place than manner changes. For the Tunisians, the distribution is inverted with more manner changes. Finally in Romanian, no difference was observed. Despite this heterogeneity across language groups, 7/13 children (no data for two Dutch participants) exhibited the same frequency for both categories. The majority (6/7) produced more manner than place changes even if the difference was not statistically significant. This preference for manner changes is confirmed by two children with more manner than place changes. Only four children used more place changes. Finally the same cross language difference was revealed in these analyses: Tunisian children used statistically more manner changes than French children. The difference between Tunisian and Dutch was not evaluated, as the Dutch group was only composed of one child. Concerning the developmental trajectory from *babbling* to *later word* periods, this domain presented the highest intra and interindividual differences. Manner changes decreased in Tunisian and Dutch. They tended to decrease in Romanian and increased in French. No general trends in use of place and manner variegation across syllables could be ascertained. For three children the percentage decreased, for three it increased and for six it stayed the same. No data was available for three other children. The same heterogeneity emerged from the changes from the *early word* to *later word* periods: French and Romanian (one subject) children used less manner changes whereas Dutch and Tunisian used more during the last period. No real trend could be identified: four children used less manner changes, four used more and two used the same frequency. Two children had no data for the *later word* period and three had no data for the period.

Discussion

The ESF-EUROCORES-OMLL (*Origin of Man, Language and Languages*) programme provided programmatic support for contemporary multidisciplinary research designed to address the seminal evolutionary perspective proposed by Charles Darwin in the mid 19th century. Diverse research cultures encompassing 'Language and the Brain', 'Language and Archaeology', 'Language and Genetics', 'Language Acquisition and Language Universals', 'Language and Animal Communication', and 'Language Evolution and Computer Modeling' were implemented within the scope of OMLL for asking critical questions about the phylogenetic origins of modern language capacities.

Through the multidisciplinary scholarly inquiries enabled by the OMLL initiative, research paradigms centered on the ontogenetic process of contemporary infants in acquiring mature human speech production capacities have been

considered a relevant contributing perspective. Central questions enabled by examination of ontogeny in human children can be understood in the context of emergence of complexity in service of increase in function, a central tenet of Darwinian evolution. Consideration of the varied parameters of the speech acquisition process enables a short time scale view of the long time scale evolutionary process that can be observed in modern language human capacities.

Issues that we have evaluated in this cross language research program are tailored to consider the emergence of complexity for human speech capacities as they are implemented in service of increasing message complexity in ontogeny. Collection of substantial databases enables statistical analysis of emerging complexity in speech patterning to enhance findings of previous small scale studies that relied on descriptive paradigms. Use of a consistent protocol for data collection and analysis across an extended period of development in diverse language environments enables a broader picture of emergence of speech complexity in early ontogeny than previous studies of single languages or smaller scale studies of pairs of languages. As such, this database can provide a source for considering at a more general level the timing of emergence of complexity in ontogeny.

As well, these comprehensive databases centered in ontogeny enable consideration of aspects of modern languages that are consistent with early emerging patterns versus aspects which reflect diverse types of phonological complexity. As we have noted, aspects of modern languages that are consistent with those observed in ontogeny, such as the use of open syllables across languages, can be seen to reflect extremely basic aspects of operation of the speech production apparatus retained in adult speakers in the face of pressure to implement diverse uses of the production system. As such, they seem to be fruitful areas of inquiry in consideration of patterns characteristic of early speakers in the formative history of communicatively oriented speech production. In turn, the vast panoply of phonological diversity characteristic of modern languages illustrates the ways in which the evolutionary process resulted in diversification of sound patterns as pressures for communicative complexity grew. These diverse phonological systems enable comparative analysis of the timing of learning to reproduce precise ambient language regularities by modern infants. Elements of children's early speech output patterns which match ambient language specific patterns point toward fruitful areas of inquiry for modeling studies in understanding the path to emergence of speech production complexity.

What did we find in this investigation of the process and products of early speech ontogeny across varied languages? We observed strong patterns of continuity between vocal patterns in the *babbling* and *early word* periods of development across the languages studied. As has been found in earlier studies of children in varied language environments, these children in four diverse language groups

exhibited a “motor core” of common production patterns (Locke, 1983; MacNeilage, Davis, Kinney & Matyear, 2000).

Concerning manner of articulation, all children used more stops and glides during the entire three periods covered by the study. Almost all children used more labials and coronals than other places of articulation from *babbling* to *later word* periods. Concerning vowels, all children except one in the *early word* period and one in the *later word* period preferred to produce LLQ vowels. Relative to intersyllabic serial trends across syllables, fewer of the children showed significant use of height changes for vowels and manner changes for consonants predicted by the Frame-Content Theory (MacNeilage & Davis, 1990). While results for these two indices of jaw predominance over independent movements of articulators were not significant in most cases, the trend for both was for a predominance of predicted patterns. This result across languages confirms at a general but not universal level, the importance of understanding production system characteristics underlying observable vocal output patterns. The lack of significant support in later developmental periods can be traced potentially to lexical effects, where children may have favored lexical forms that are produced frequently. To understand this issue comprehensively, the interface between lexical patterns and phonetic patterns should be examined more fully in this cohort.

Only in the *later word* period, when the children were producing a much larger number of meaningful words, was there an explosion of phonetic diversification signaling emergence of complexity in production system capacities. Emergence of later appearing sounds, including fricative, affricate, and liquid manners of articulation and dorsal and guttural places of articulation was apparent. Expansion of the vowel space to include diverse vowel types not related to the LLQ constraint was also apparent in the *later word* period. This increase in output complexity seems largely related to more control over the speech production system enabling increase in capacities for matching language forms. Increases were not clearly related to precise ambient language patterns of input as they were consistent across languages.

Children in this study were learning different languages with diverse phonetic, phonological and grammatical systems. One more issue at hand in this study was to observe the data in a cross language perspective in order to determine how and when the typological characteristics of a language play a role on the timing and trajectory of speech production pattern acquisition. Only very few cross language differences emerged from this data. Moreover, the differences were not consistent from one stage to another. When looking at place of articulation for instance, there was a statistically significant difference between French and Romanian vs. Dutch and Tunisian in the *babbling* period. In the *early word* period, no difference was seen between languages. Finally, at the *later word* period, only French children behaved differently than Tunisian children. This variability can be explained by a

small number of children per language group and high interindividual differences among these language groups. In addition, we tried to find an explanation for these very inconsistent cross language differences by comparing the frequencies obtained by the children to those of the 1000 word dictionary data for each language which was considered representative of their ambient language input. In most cases, no clear input related explanation was possible. Overall, there was little evidence of learning to reproduce precise regularities related to sounds and sequences in input from the ambient language environment in these three earliest periods of development.

Research directions

As with many large scale projects addressing difficult issues, we have perhaps generated more questions than definitive answers. However, we have refined our ability to ask higher level questions with this comprehensive database and longitudinal approach to understanding the emergence of speech production complexity. Several areas of future research with this corpus should provide fruitful input to understanding the emergence of complexity in this dimension of language capacities more fully. As these analyses are completed, we can come closer to the goal of considering the utility of child speech ontogeny for understanding the process of phylogenetic change that has resulted in modern human language capacities.

Crucially, to achieve the goal of considering phylogeny from the perspective of ontogeny, child to ambient language comparisons should be completed at a more definitive level that was possible in this initial data analysis. That project is in progress, supported by more powerful statistical evaluation tools to understand the timing and types of ambient language matches in children both within and across the four languages.

An additional level of analysis for language input values in considering emergence of ambient language regularities lies in consideration of phonological properties of child directed speech style in input to children. Child directed speech has been shown to display diversity from adult-directed speech in a number of important dimensions (Fernald & Morikawa, 1993; Newport, Gleitman & Gleitman, 1977). Relevant to the phonological input component, a recent analysis of Korean (Lee, Davis & MacNeilage, 2008) has shown statistically significant differences in frequency of child directed speech input values for Korean ambient language features from those found in Korean adult directed speech input. This result suggests that mothers may implement differences from adult phonological regularities to highlight ambient language patterns in speech to their infants in early periods of acquisition. Infant directed speech input should be investigated in these cross language corpora to characterize input values to young children more validly. From

another perspective, language to language comparisons of phonological types and frequency of occurrence of those types will provide an additional level of needed information for fully understanding the nature and pressures for emergence of ambient language precision in the phonological dimension.

One major area of diversity from the patterns observed with the present level of macro-analysis in speech patterns of these children in four languages was the level of child individual differences from dominant group patterns. For all of the indices analyzed at all three developmental periods, individual children produced patterns that differed from the dominant group trends found for their own language as well as, in a few cases, dominant trends that were found to be characteristic across the four languages. Clearly, a micro-level of analysis of individual children within and across languages needs to be added to the present macro-analysis represented in this chapter reporting our programmatic progress to date. Child individual differences and when they begin to appear can help to understand when and how cognitive development may become a driving force in early phonological development (see Vihman & Croft, 2007, for a statement of this position) relative to the production system oriented hypotheses that have been explored here.

Another fruitful area for extension of this research program should include consideration of typologically more diverse languages than the sample represented in this corpus. Languages such as Navajo or Hawaiian represent major cultural and genetic differences that are lacking in these languages. As well, more diverse language families could enable analysis of a wider spectrum of sound and sequence properties. This diversity could help to consider the nature-nurture issue at the heart of understanding the nature of ontogeny of modern language more fully.

The goal of this program of research into ontogeny of early speech production capacities in modern infants was to provide a finely grained level of understanding of the process of acquisition of the phonological component of modern language. Phonology lies at the intersection of complex knowledge and complex behaviour. This aspect of language function should be of seminal interest to gaining insight into the evolutionary process of human language. Of necessity at its onset and in early periods, early evolution of language capacities was embodied in intersection of emerging knowledge based on perceptual access with emerging capacities for planning and producing complex vocal output. This input-knowledge-output continuum would have been in service of communicating needed messages for functioning in an *early* cultural context. In the same way, the input-knowledge-output continuum observable in modern infants from its simplest onset in the first year of life illustrates emergence of a complex system in service of communicating needed messages for function in a *modern* cultural context.

Acknowledgements

This work was supported by the EUROCORES Programme “The Origin of Man, Language and Languages” (OMLL), the French CNRS program “Origine de l’Homme, du Langage et des Langues” (OHLL), and research grant # HD 27733–10 from the U.S. Public Health Service.

Project participants were: Sophie Kern (project leader) & Laetitia Savot (research assistant), Laboratory Dynamique du Langage, Lyon, France, Inge Zink (principal investigator), Mieke Breuls & Annemie Van Gijssel (research assistants), Exp. ORL, Dept. Neurosciences, K.U Leuven, Belgium, Barbara Davis, Peter MacNeilage & Chris Matyear, Speech Production Laboratory, Austin, Texas, USA.

References

- Bickley, C. (1983). Acoustic evidence for phonological development of vowels in young infants. Paper presented at the 10th Congress of Phonetic Sciences, Utrecht.
- Boliek, T., Hixon, P., Watson, W., Morgan (1997). Vocalization and breathing during the second and third years of life. *Journal of Voice*, Volume 11, Issue 4, 373–390.
- Buhr, R. D. (1980). The emergence of vowels in an infant. *Journal of Speech and Hearing Research*, 12, 73–94.
- Davis, B. L., & MacNeilage, P. F. (1995). The Articulatory Basis of Babbling. *Journal of Speech and Hearing Research*, 38, 1199–1211.
- Davis, B. L., & MacNeilage, P. F. (2000). An embodiment perspective on the acquisition of speech perception. *Phonetica*, 57(Special Issue), 229–241.
- Davis, B. L., MacNeilage, P. F., & Matyear, C. L. (2002). Acquisition of Serial Complexity in Speech Production: A Comparison of Phonetic and Phonological Approaches to First Word Production. *Phonetica*, 59, 75–107.
- de Boysson-Bardies, B., Sagart, L., & Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language*, 11(1), 1–15.
- de Boysson-Bardies, B., Hallé, P., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, 16, 1–17.
- de Boysson-Bardies, B., Vihman, M. M., Roug-Hellichius, L., Durand, C., Landberg, I., & Arao, F. (1992). Evidence of infant selection from target language: A crosslinguistic phonetic study. In C. A. Ferguson & L. Menn & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications*. Monoton, MD: York Press.
- de Boysson-Bardies, B. (1993). Ontogeny of language-specific syllabic production. In B. de Boysson-Bardies & S. de Schoen & P. Jusczyk & P. F. MacNeilage & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 353–363). Dordrecht: Kluwer Academic Publishers.
- Fenson, L., Dale, P., Reznick, S., Thal, D., Bates, E., Hartung, J., Tethick, S., & Reilly, J. (1993). *MacArthur Communicative Development Inventories: User’s guide and technical manual*. San Diego: CA Singular Publishing Group.

- Fernald, A. (1992). Maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J.H. Barkow, L. Cosmides, & J. Tooby (eds.), *The adapted mind: Evolutionary Psychology and the Generation of Culture* (pp. 391–428). Oxford: Oxford University Press.
- Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development* 64(3), 637–656.
- Gildersleeve-Neumann, C., & Davis, B. L. (1998). Production versus ambient language influences on speech development in Quichua. Paper presented at the Annual Meeting of the American Speech, Hearing and Language Association, San Antonio, Texas.
- Green, J.R., Moore, C.A., & Reilly, K.J. (2002). The sequential development of jaw and lip control for speech. *Journal of Speech, Language, and Hearing Research*, 45, 66–79.
- Kent, R. D., & Bauer, H. R. (1985). Vocalizations of one-year olds. *Journal of Child Language*, 12, 491–526.
- Kern, S. & Davis, B.L. (2009, in press). Emergent complexity in early vocal acquisition: Cross-linguistic comparisons of canonical babbling. In Chirotan, I. Coupé, C., Marsico, E. & Pellegriano, F. *Approaches to Phonological Complexity*, Benjamins.
- Kopkalli-Yavuz, H., & Topbas, S. (2000). Children's preferences in early phonological acquisition: How does it reflect sensitivity to the ambient language? In A. Göksel & C. Kerslake (Eds.), *Studies on Turkish and Turkic Languages* (pp. 289–295). Wiesbaden: Harrassowitz.
- Ladefoged, P. & Maddieson, I. (1996). *The Sounds of the World's Languages*, Oxford, UK: Blackwell Publishers.
- Lee, S., Davis, B. L. & MacNeilage, P.F. (2008). On the phonetic characteristics of input speech: A study of Korean, *Journal of Child Language*.
- Lee, S., Davis, B.L. & MacNeilage, P.F. (2007). Frame Dominance' and the Serial Organization of Babbling, and First Words in Korean-Learning Infants, *Phonetica*, 64, 217–236.
- Levitt, A. G., & Utman, J. G. A. (1992). From babbling towards the sound systems of English and French – a longitudinal 2-case study. *Journal of Child Language*, 19, 19–49.
- Lieberman, P. (1980). On the development of vowel production in young children. In G. H. Yeni-Komshian & J. F. Kavanagh & C. A. Ferguson (Eds.), *Child phonology 1: Production*. New York, NY: Academic Press.
- Locke, J. L. (1983). *Phonological acquisition and change*. New York, NY: Academic Press.
- MacNeilage, P. F., & Davis, B. L. (1990). Acquisition of Speech Production: Frames, Then Content. In M. Jeannerod (Ed.), *Attention and Performance XIII: Motor Representation and Control* (pp. 453–476). Hills: Lawrence Erlbaum.
- MacNeilage, P. F., & Davis, B. L. (1993). Motor explanations of babbling and early speech patterns. In B. de Boysson-Bardies & S. de Schonen & P. Jusczyk & P. F. MacNeilage & J. Morton (Eds.), *Changes in Speech and Face Processing in Infancy: A Glimpse at Developmental Mechanisms of Cognition*. Dordrecht: Kluwer.
- MacNeilage, P. F., & Davis, B. L. (2000). On the Origin of Internal Structure of Word Forms. *Science*, 288, 527–531.
- MacNeilage, P. F., Davis, B. L., Kinney, A., & Matyear, C. L. (2000). The Motor Core of Speech: A Comparison of Serial Organization Patterns in Infants and Languages. *Child Development*, 2000(1), 153–163.
- Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.
- Newport, E., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself: some effects and non-effects of maternal speech style. In C. Snow, & C. A. Ferguson (eds.), *Talking to*

- Children: Language Input and Acquisition* (pp. 109–150). Cambridge, MA: Cambridge Press.
- Oller, D. K., Wieman, L. A., Doyle, W. J., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language*, 3, 1–11.
- Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian & J. F. Kavanagh & C. A. Ferguson (Eds.), *Child phonology 1: Production*. New York, NY: Academic Press.
- Oller, D. K., & Eilers, R. E. (1982). Similarity of babbling in Spanish- and English-learning babies. *Journal of Child Language*, 9, 565–577.
- Oller, D. K., & Steffans, M. L. (1993). Syllables and segments in infant vocalizations and young child speech. In M. Yavas (Ed.), *First and second language phonology* (pp. 45–62). San Diego: Singular Publishing Co.
- Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Oller, D.K. & R. Delgado, (1990). *Logical international phonetic programs*. Miami: Intelligent Hearing Systems.
- Roug, L., Landberg, I., & Lundberg, L. (1989). Phonetic development in early infancy: A study of four Swedish children during the first eighteen months of life. *Journal of Child Language*, 16, 19–40.
- Rousset, I. (2004). Structures syllabiques et lexicales des langues du monde: Données, typologies, tendances universelles et contraintes substantielles. Unpublished doctoral dissertation, Université Stendhal, Grenoble, France.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8 Month Old Infants. *Science*, 274, 1926–1928.
- Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian & J. F. Kavanagh & C. A. Ferguson (Eds.), *Child phonology 1: Production*. New York, NY: Academic Press.
- Stoel-Gammon, C. (1985). Phonetic inventories 15–24 months – a longitudinal study. *Journal of Speech and Hearing Research*, 28, 505–512.
- Stoel-Gammon, C., & Harrington, P. (1990). Vowel systems of normally developing and phonologically disordered children. *Clinical Linguistics and Phonetics*, 4, 145–160.
- Stoel-Gammon, C., & Cooper, J. (1984). Patterns of early lexical and phonological development. *Journal of Child Language*, 11(2), 247–271.
- Teixeira, E. R., & Davis, B. L. (2002). Early sound patterns in the speech of two Brazilian Portuguese speakers. *Language and Speech*, 45(2), 179–204.
- Tyler, A. A., & Langsdale, T. E. (1996). Consonant-vowel interaction in early phonological development. *First Language*, 16, 159–191.
- Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment to the continuity issue. *Language*, 61, 397–445.
- Vihman, M. M., Ferguson, C. A., & Elbert, M. F. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics*, 7, 3–40.
- Vihman, M. M. (1992). Early syllables and the construction of phonology. In C. A. Ferguson & L. Menn & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications*. Monckton, MD: York Press.
- Vihman, M. M. (1996). *Phonological Development: The Origins of Language in the Child*. Cambridge, Massachusetts: Blackwell Publishers.

- Vihman, M. M. & Croft, W. (2007). *Phonological development: Toward a 'radical' templatic phonology*. *Linguistics*, 45, 683–725.
- Werker, J. F., & Tees, R. C. (1984). Cross-Language Speech Perception: Evidence for Perceptual Reorganization during the First Year of Life. *Infant Behavior and Development*, 7, 49–63.
- Werker, J. F., & Lalonde, C. E. (1988). Cross language speech perception: initial capabilities and developmental change. *Developmental Psychology*, 24, 672–683.
- Werker, J.F., & Curtin, S. (2005). PRIMIR: A developmental model of speech processing. *Language Learning and Development*, 1(2), 197–234.
- Werker, J.F., & Tees, R.C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, 46(3), 233–251.

Language and animal communication

The primate roots of human language

Primate vocal behaviour and cognition in the wild

Klaus Zuberbühler^{1, 2}, Karim Ouattara^{2, 3}, Anderson Bitty²,
Alban Lemasson^{2, 3} and Ronald Noë^{2, 4}

Our research describes the natural communication in different primates, as well as their underlying cognitive abilities. We are interested in the communication skills and social cognition of monkeys and apes in their natural habitats. For this purpose, we observe wild groups in Africa and Asia during their daily activities in order to understand what communication signals they are able to produce, under what circumstances they produce them, and what sorts of responses they elicit from listeners. Once we have gathered enough information to suspect a relationship between a particular call type, a set of events and a typical response, we conduct a field playback experiment to determine the call's meaning. During such an experiment we play back a recorded example of a particular call to a naïve receiver in order to study its response. For example, chimpanzees produce different types of food grunts depending on the type of food they find, such as apples or bread. We found that these calls were indeed meaningful to other chimpanzees. For instance, if we played back recordings of 'apple grunts' listeners were more likely to look for food in places where they previously found apples, but not bread, and vice versa. In another study, we found that some monkeys are able to combine some of their vocalizations in systematic ways to create different combinations with separate meanings. Our playback experiments showed that it was the combinations of calls, rather than the individual calls themselves, which carried the meaning, an example of simple primate 'grammar'.

-
1. School of Psychology, University of St Andrews, St Andrews, Scotland UK.
 2. Centre Suisse de Recherches Scientifiques, Tai Monkey Project, Abidjan, Ivory Coast.
 3. Université de Rennes 1, Ethos – Ethologie animale et humaine, Paimpont, France.
 4. Ethologie des Primates, CNRS & Université de Strasbourg, Strasbourg, France.

Introduction

The comparative approach is one major way to study human language evolution with non-human primates playing a key role. By studying the communicative abilities of our closest living relatives, and particularly their vocal behaviour, it is possible to recognise the evolutionary origins of the various anatomical, neural, behavioural and cognitive features that underlie human language. A functional description of a species' vocal repertoire often provides a first step in any such study. For primates, such analyses typically reveal a finite number of acoustically distinct sounds given by different group members in more or less specific contexts (e.g. Range & Fischer 2004; Lemasson & Hausberger 2004). Amongst all primates, humans are clearly an exceptionally vocal species. From early on, human infants produce a rich array of sounds, a manifestation of their potent urge to engage in social activities, obtain goods, or affect others' behaviours. For adults, the International Phonetic Alphabet (IPA) lists over 100 different phones to describe the perceptually meaningful speech sounds of the world's 6,000 or so languages. Most IPA listed sounds can be produced as single utterances, but some others do not have their own acoustic identity. Human vocal behaviour is also unusual in that sound utterances are commonly combined into rapid sequences to form higher order units that carry meaning.

These two aspects of vocal production, i.e. richness of the acoustic portfolio and predisposition to combine basic acoustic units into more complex utterances, appear to be uniquely human traits, calling for an evolutionary investigation. What are the phylogenetic roots of these behaviours? In line with our commitment to the comparative approach, we review a number of studies on primate vocal behaviour, first dealing with aspects of vocal production, and then attending to matters of perception and understanding.

Origins of articulation

Most researchers agree that non-human primates have little voluntary control over their articulators. In particular, there has been no good evidence of any primate combining basic call types into more complex units, no cases of invention of new calls or voluntary modification of existing ones (Tomasello & Zuberbühler 2002). For example, cross-fostered macaques (*Macaca mulatta*, *M. fuscata*) retain their species-specific vocal repertoire and show virtually no changes in the acoustic structure of their calls, despite ample and exclusive exposure to the other species' vocalisations (Owren et al. 1993). Hybrid gibbons (*Hylobates lar* x *pileatus*) produce songs with an intermediate acoustic structure compared to those of the two

parent species (e.g. Geissmann 1984). To some degree, such findings can probably be explained with species-specific anatomical differences in the vocal tract, which will constrain the acoustic structure of the signals an individual can produce. However, non-human primates also appear to be largely unable to control sound production at the level of larynx, and to adjust the shape of their vocal tracts to compensate for any anatomical differences. For example, despite considerable training effort, it was not possible to get rhesus macaques to alter the acoustic features of their vocalizations (Sutton et al. 1973), while squirrel monkeys (*Saimiri sciureus*) have been reported to produce most call types soon after birth, even if reared in isolation (e.g. Winter et al. 1973; Hammerschmidt et al. 2001). From an evolutionary perspective, these findings are bewildering. Why are humans so different in these crucial aspects, given the many parallels with non-human primates in the general anatomy and social behaviour? More specifically, what are the phylogenetic roots of our elaborate vocal tract control, and which evolutionary events might have been responsible for this?

Ontogenetic flexibility

A number of studies have revealed some complexities with regards to the notion of non-human primates' vocal rigidity. Some researchers have reported ontogenetic modifications in call structure that clearly go beyond mere maturational effects. An important finding is that repertoire changes generally tend to occur after significant social events. For example, after regrouping pygmy marmosets (*Cebuella pygmaea*) have been observed to change the acoustic structure of their 'trill' calls (Snowdon & Elowson 1999), while in cotton-top tamarins (*Saguinus oedipus*) changes in social rank can affect the acoustic structure of 'chirp' calls (Roush & Snowdon 1999). Social factors also affect the acoustic structure of 'pant hoot' vocalisations in chimpanzees. Observations at various zoos have shown that individuals can converge on structurally unique local pant hoot variants, suggesting that some sort of learning is involved in call production (Marshall et al. 1999).

Work with Campbell's monkeys (*Cercopithecus campbelli*) has produced a detailed picture of some of the underlying mechanisms of call modifications during adulthood. Using a telemetric recording technique, eight female captive Campbell's monkeys, housed in a natural one-male social group at the 'Station Biologique de Paimpont' (France), were monitored in terms of their vocal behaviour. Results showed that their repertoire contained three levels of organisation. First, it was possible to describe ten basic call types of different functional categories (e.g. alarm threat, distress, contact). Four of these call types contained significant acoustic variation, and it was thus possible to further discriminate them into specific *sub-types*. One of these four call types, the combined-harmonic call (CH), serves to

maintain contact in the monkeys' naturally dense forest habitat and is frequently exchanged by individuals during affiliative interactions, with individuals responding to one another. Detailed acoustic analyses of the CH call then revealed significant acoustic variations, which could be categorised in terms of six distinct sub-types (CH1–6). The production of these sub-types was context-dependent, rather than a product of individual differences or other factors (Lemasson 2003; Lemasson et al. 2004).

One of them, the CH6 subtype, could be further discriminated into different *variants* (Lemasson 2003), and it was found that each individual produced between one and four relatively stereotyped variants, differing in the shape of their frequency modulations (Lemasson et al. 2003). The surprising finding was that some of these variants were shared by certain individuals. Variants were not context-related since females frequently responded with dissimilar variants to each other in a given context. Analyses over a three-year period revealed that if two individuals shared a particular variant, this was either because they were converging towards a novel variant or because one was converging towards another individual's existing variant. Proximity in age or rank did not explain these effects, nor did kinship patterns as 63% of sharing occurred between unrelated females. Instead, an individual's degree of vocal sharing was correlated with its level of social affiliation, as measured notably in terms of play behaviour and vocal exchange patterns (Lemasson & Hausberger 2004). The number of affiliative interactions between two females, in other words, appeared to be the best predictor of the propensity to share a variant. Campbell's monkey females form long-term stable bonds (Lemasson et al. 2006) and socially isolated animals were less likely to share variants but produced variants that were dissimilar to those of other group members (Lemasson et al. 2003; Lemasson & Hausberger 2004). Vocal sharing was also observed in two other captive groups (Lemasson & Hausberger 2004). An ongoing study with free-ranging Campbell's monkeys in Tai forest is currently investigating this phenomenon in the wild (Ouattara et al. 2009, in press).

Although vocal sharing has been observed in some animal taxa (birds: Hausberger et al. 1995, Brown & Farabaugh 1997; cetaceans: Janik 2000; humans: Milroy 1987) it has never been described in non-human primates. For Campbell's monkeys, the general conclusion was that the vocal repertoires of these females remained plastic, at least at the level of the call variant. Throughout their adult life, different variants appeared, disappeared or underwent modification over time, typically in response to changes in the current social relationships (Lemasson & Hausberger 2004). In playback experiments, females could distinguish between different call variants of familiar group members, suggesting that they formed part of a long-term social memory (Lemasson et al. 2005). While playbacks of some of the females' currently produced call variants reliably triggered vocal responses

from other group members, the same females' no longer used variants (recorded 4 years earlier) never did, similar to the calls of unfamiliar females. It is possible that this vocal sharing is used by these primates as a sort of 'social badge' to advertise their current social bonds to other group members. In line with this hypothesis, there were drastic changes in the patterns of vocal sharing after the replacement of the single adult male (Lemasson & Hausberger 2004), who plays a key role in the group's social life (Lemasson et al. 2006).

In sum, although much of the current theory suggests that non-human primates have fixed and immutable calls, closer examination often reveals that acoustic modification can occur within certain call types, even during an animal's adult lifetime. These changes are not simply maturational by-products of changes in body size, or the like, but frequently related to the individual's social environment.

Mechanisms of articulation

The richness of the human vocal repertoire is the result of a highly adapted vocal tract, consisting of various articulators that can perform a number of controlled movements: lips, tongue, mandible, velum, and larynx. A particularly useful theoretical foundation for studies of primate communication has been the source-filter theory developed for speech research (Fant 1960). The speech signal is conceptualised as sound generated by a source, the oscillating vocal folds of the larynx, and its vibration periodicity determines the fundamental frequency of the utterance. Source sound then enters the supra-laryngeal vocal tract, which has a number of resonances and acts as a bank of band-pass filters, allowing only narrow ranges of frequencies to pass. The geometry of the human vocal tract is adjustable due to several flexible articulators. The resulting patterns of vocal tract resonances, along with the spectral peaks they produce in the vocal signal, have been termed 'formants'. Calls produced by sources other than laryngeal vibrations also occur, but their acoustics has not been studied well.

More recently, the source-filter theory has been applied to explain the acoustic structure of animal calls (Lieberman et al. 1969; Nowicki 1987; Owren & Bernacki 1988; Riede & Fitch 1999). The formant patterns in animal vocalisations, including those of non-human primates, are often simple compared to those in human speech (Lieberman et al. 1969; Owren et al. 1997). Because they are typically evenly spaced, they are thought to be the product of a vocal tract that is shaped like a uniform tube (Fitch 1997; Riede & Fitch 1999). It has long been thought that the reason for the human-animal difference in vocal structure stems from the fact that the resting position of the human larynx is much lower compared to that of non-human primates or very young children (Negus 1949). A short larynx is likely to block any articulatory movements of the tongue, providing an explanation for the meagre

acoustic flexibility (Lieberman et al. 1969). Cranial reconstructions of early hominids, particularly *Homo erectus*, have shown that the vocal tract anatomy of these individuals was more likely to resemble that of modern chimpanzees, suggesting that these individuals were unable to produce the entire range of modern speech sounds (e.g. Lieberman et al. 1998). Because of the inflexible vocal tract, it was thought that acoustic variation in non-primates was mainly generated at the level of the larynx, for example by adjusting the pressure of airflow from the lungs.

However, recent empirical work has generated a more complicated picture. Rhesus macaques clearly engage in articulation during vocal behaviour, particularly with their lips and mandibles (Hauser et al. 1993). Other studies have shown that macaques can control the resonance properties of the sounds independently of the laryngeal source and hence must be capable of vocal tract filtering (Hauser & Schön-Ybarra 1994). Using cineradiography, Fitch (2000) has observed the movements of the articulators in live animals during vocalisation, including those of cotton-top tamarins, *Saguinus oedipus*. He found that the larynx could be retracted and lowered substantially during vocalisation from a high resting position in the naso-pharynx, with the velum closing the nasal passage. Second, during chewing and swallowing tongue movement was highly flexible, suggesting that individuals have substantial control over their tongues. These two findings have cast doubt on the idea that the simple acoustic patterns observed in non-human primates are the result of a high larynx, which obstructs tongue movement (e.g. Riede et al. 2005, 2006). Primate tongues are able to perform complex manoeuvres, particularly when the larynx is lowered during vocalisation.

Semantic encoding

Recent work on Diana monkeys has produced more evidence of active vocal tract filtering in non-human primates (Riede & Zuberbühler 2003a&b). When wild adult male Diana monkeys (*Cercopithecus diana*) of Taï forest, Ivory Coast, encounter a crowned eagle or a leopard, two of their natural predators, the adult male produces alarm calls to warn others and to signal detection to the predator (Zuberbühler et al. 1997). Acoustic analyses have shown that the alarm calls to the two predator types differ slightly in their acoustic fine structure, the result of active vocal tract filtering (Zuberbühler 2000a, Riede & Zuberbühler 2003a, Figure 1).

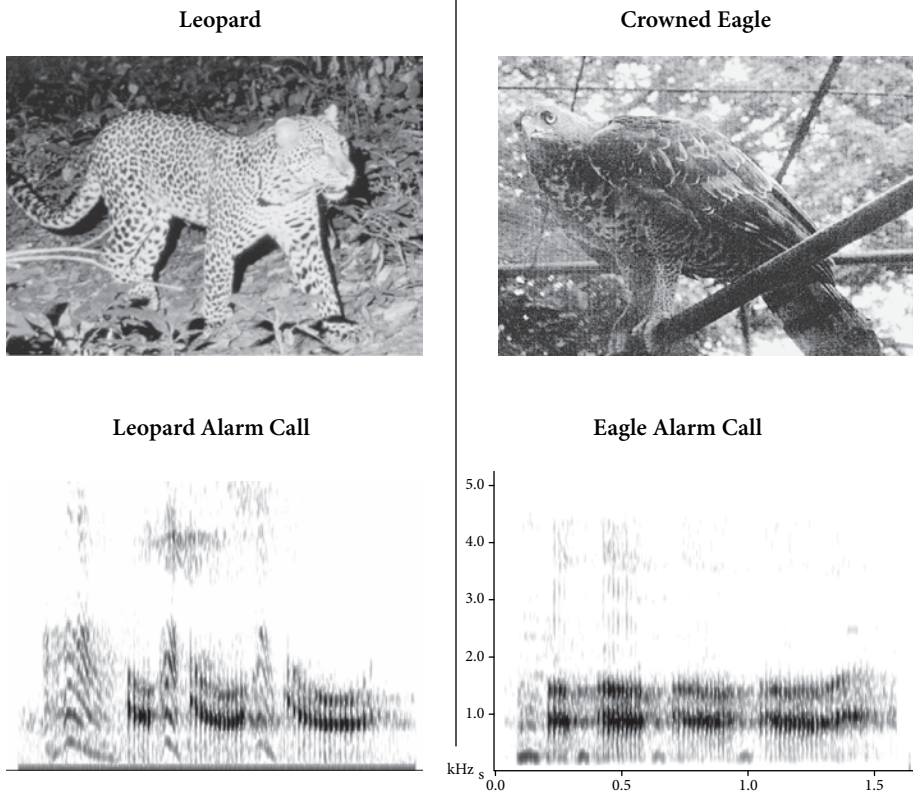


Figure 1. Male Diana monkey alarm calls to a crowned eagle (a) and a leopard (b). The acoustic fine structure of these alarm calls is predator-specific, with leopard alarm calls featuring a downward transition at the beginning of the calls. Photo leopard: David Jenny; photo crowned eagle: Klaus Zuberbühler

Figure 1b shows that the two call types differ most prominently in the frequency modulation of the first formant. Leopard alarm calls exhibit a three-fold stronger downward modulation than eagle alarm calls. Anatomical data have shown that Diana monkey vocal tracts do not resemble a uniform tube but show flexible constrictions at various sites, the foundation for vocal tract filtering. These anatomic data, combined with acoustic modelling, revealed that a combination of lip aperture, mandible movement and larynx lowering was able to closely replicate the vocal patterns produced by wild animals (Riede et al. 2005). Non-human primates, in other words, have some control over their articulators, and may use this ability to transmit semantic information (see Ouattara et al. 2009 for a recent example). The question then is why humans are so much better at controlling their articulators

and why vocal behaviour in non-human primates appears to be so tightly bound to specific psychological states.

Work on human speech disorders has provided some insights concerning the vastly superior articulatory abilities in humans. This research has shown that damage to a particular area in chromosome 7, the FoxP2 gene, can lead to language disorders in which individuals experience substantial difficulties in producing coherent and fluent speech, while other aspects of their cognition remain unaffected (Lai et al. 2001). The phylogenetic history of the FoxP2 gene is particularly interesting. Enard et al. (2002) have argued that as little as 200,000 years ago two mutations in FoxP2 became stabilized in the hominid population ancestral to ours. In other words, the current form of the FoxP2 gene, crucial for speech production, has only been around for 10,000 or so generations. One remarkable conclusion from these findings is that, prior to this event, our ancestors did not have the articulatory abilities of modern humans, perhaps more resembling today's non-human primates in their vocal behaviour.

Production of syntactic cues

Human languages can be described as collections of syntactic rules. They allow speakers to construct an infinite number of previously unheard messages or "to make infinite use of finite means" (von Humboldt 1836). Animal communication, in contrast, is event-bound with no comparable evidence of creativity (Ghazanfar & Hauser 1999), despite the fact that naturally occurring combinatorial properties have been reported in animal communication systems (e.g. Hailman & Ficken 1987). Recent work with free-ranging putty-nosed monkeys (*C. nictitans*) in Nigeria has shown that this simple dichotomy is wrong and that some primates can assemble different call types into more complex call sequences, which then serve as the carriers of meaning (Arnold & Zuberbühler 2006a, b; Arnold et al. 2008a). Male putty-nosed monkeys produce two main loud call types, the pyows and hacks, which they give in response to a number of events, typically some type of disturbance including the presence of key predators, such as crowned eagles or leopards. However, the two call types do not function as predator-specific alarm calls because both call types can occur in response to both predators. Yet some context-specificity does emerge if the unit of analysis is shifted from individual calls to call sequences (Figure 2).

Another important finding was that the male putty-nosed monkeys combined the two basic call types in a specific way before they wished to travel. In this particular context, males usually begin calling with a few pyows, followed by a few hacks, which generates a perceptually distinct pyow-hack sequence. These pyow-hack combinations can be given independently or as part of other call sequences,

Call position	N	1	2	3	4	5	6	7	8	9	10	11	12
Eagle trials													
Pure hack	10	H	H	H	H	H	H	H	H	H	H	H	H
Transitional	5	H	H	H	H	H		P	P	P	P	P	P
Pure pyow	2	P	P	P	P	P	P	P	P	P	P	P	P
Pure hack-PHS _{ins}	2	H	H	H	H	H		P	H	H		H	H
Transitional-PHS _{ins}	1	H	H	H	H	H		P	H	H		P	P
Leopard trials													
Pure pyow	5	P	P	P	P	P	P	P	P	P	P	P	P
PHS	5	P	P	H	H	H							
Pure pyow-PHS _{start}	4	P	P	H	H	H	P	P	P	P	P	P	P
PHS – PHS	1	P	H	H			P	H					
Transitional	1	H	H			P							

Figure 2. Call sequences consisting of two different alarm calls, pyows and hacks, produced by free-ranging putty-nosed monkeys in response to crowned eagle or leopard. H=hack, P=pyow; PHS=pyow-hack sequence (Arnold et al. 2008)

usually separated by a perceptible break (Arnold & Zuberbühler 2006; Arnold et al. 2008). Specifically, they can be given as part of an anti-predator response or as part of normal daily routines. In one playback experiment, call sequences were triggered from males by simulating the presence of a leopard. If the male's calling response contained a pyow-hack series the group travelled significantly further in the subsequent 30min than if the male called but did not produce a pyow-hack series (Arnold & Zuberbühler 2006; Figure 3).

Our ongoing research in the Tai Forest of Ivory Coast suggests that similar patterns are likely to emerge from free-ranging Campbell's monkeys (Zuberbühler 2000b; 2002; Ouattara et al. 2009, in press). Male Campbell's monkeys sometimes produce a pair of brief and low-pitched "boom" vocalizations before giving a series of alarm calls about 30 seconds later. "Boom" -*introduced* alarm call series are given to a number of disturbances, such as a falling tree or large breaking branch, or the far-away alarm calls of a neighbouring group. Common to these contexts is the lack of immediate danger, unlike when the caller is surprised by a predator.

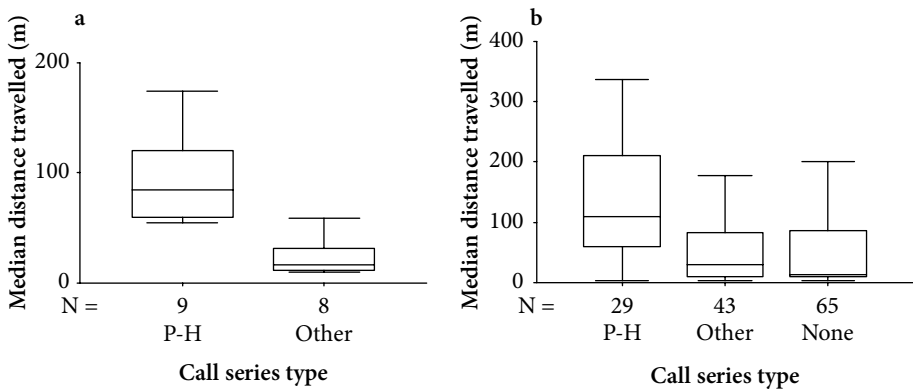


Figure 3. Locomotor behaviour of putty-nosed monkey groups in response to different male alarm call sequences (a) experimental data; (b) observational data (Arnold & Zuberbühler 2007)

The subsequent alarm call series produced by male Campbell's monkeys consist of an acoustically highly heterogeneous class. With careful analyses it is possible to distinguish at least five different call types, in addition to the “boom” calls that introduce them. Results show that callers assemble these calls into context-specific sequences and our ongoing research is in the process of elucidating the relationships between call context and sequence composition (Ouattara et al. in prep.).

Black-and-white colobus monkeys (*Colobus polykomos* and *C. guereza*), finally, are another example of how meaning can be encoded by arranging individual vocal units in specific ways. These monkeys produce only one main call type, the roar, in response to various disturbances. However, callers generate context-specific acoustic differences by assembling the individual roars into roaring sequences that differ in the number of roars. Roaring sequences to leopards are uniformly short, typically consisting of only a small number of roars, whereas sequences to crowned eagles usually consist of a large number of roars. Detailed acoustic analyses failed to detect any relevant acoustic differences between individual roars to eagles and leopards, suggesting that these monkeys convey meaning by the temporal patterning of their calling (Schel et al. 2009).

Pragmatics of call production

There is some evidence from blue monkeys (*C. mitis*) that non-human primates take ongoing social events into account when responding to external non-social events, such as the presence of a predator. Blue monkeys are relatively common in East and South African forests and groups typically consist of one adult male and

several adult females and their offspring. After reaching adulthood, males typically leave their natal group to try and take over another group to maintain tenure for a number of years, and multi-male influxes have been observed (Cords 1988). Males thus have a strong biological interest to protect the group from predation, and to keep out rival males, during this reproductively limited time. Blue monkey males vigorously produce loud alarm calls to predators, best described as 'hacks' (or 'ka-trains') and 'pyows'. Blue monkeys are close relatives of putty-nosed monkeys and the males' calls are acoustically similar. In one study, conducted in Budongo Forest, Uganda, it was found that males produce alarm calls that are acoustically similar to those of putty-nosed monkeys: hacks to crowned eagles and pyows to leopards as well as a range of other disturbances (Papworth et al. 2008). Although this finding is indicative that blue monkeys' alarm calls function in a rudimentary referential way, it was not clear whether callers intend to inform their nearby audience about the nature of the threat (Cheney & Seyfarth 1990; Tomasello et al. 2005).

Whether or not alarm callers take their audience into account is of considerable relevance for questions concerning the origins of human language. A relevant study has been conducted by Wich & de Vries (2006), who demonstrated that male Thomas langurs (*Presbytis thomasi*) continued alarm calling to a tiger model until all adult group members had produced at least one alarm call, suggesting that callers monitored their audience while producing alarm calls.

To further investigate whether non-human primates take into account their audience when producing alarm calls we conducted a follow-up study. In free-ranging blue monkeys of Budongo Forest, we manipulated the degree of threat experienced by other group members, and monitored the males' alarm calling response. We did this by playing back series of alarm calls recorded from males that originally responded to eagles and leopards. As an indicator of threat, we measured the distance of both the caller and the nearest female to the speaker during each trial. Crowned eagles pose a severe threat to the monkeys of Budongo Forest, but leopards have become exceedingly rare and many monkeys may have no experience with this predator. Results showed that when responding to series of hacks, indicating the presence of an eagle, males responded predominantly with hacks, but produced significantly more calls if their group members were close to the playback stimulus than far away, regardless of their own position. When responding to series of pyows, indicating a range of disturbances, males responded with pyows, but call rates were independent of others' distances (Papworth et al. 2008). We thus concluded that males took into account the degree of danger experienced by other group members (Figure 4).

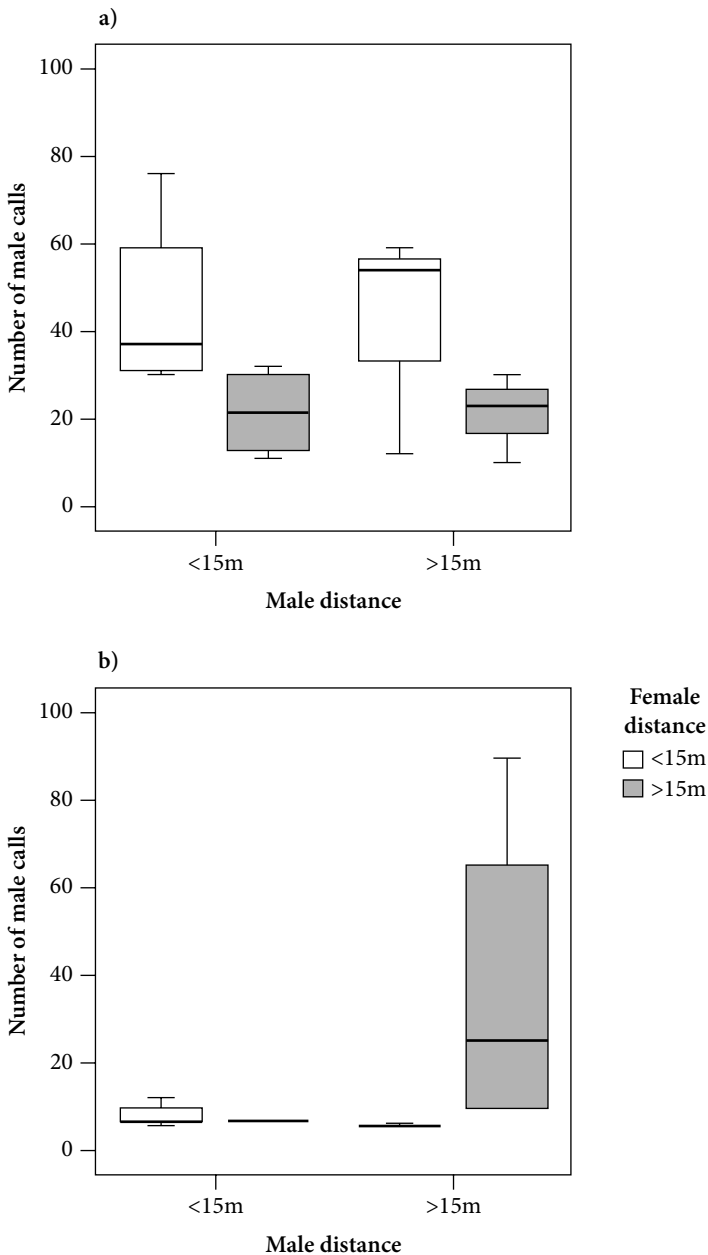


Figure 4. Vocal responses of male blue monkeys to neighbours giving series of (a) hacks, indicating the presence of a crowned eagle, (b) pyows, indicating the presence of an unspecified terrestrial disturbance. Male calling rates were not only affected by the type of disturbance but also by the distance of the group to the suspected disturbance, regardless of the caller's own distance to it (Papworth et al. 2008)

Call comprehension

If it is correct that modern humans were the first primates to possess elaborate articulatory abilities, then this poses an intriguing problem for explaining language evolution. If speech is so young, as for example suggested by Enard et al. (2002), then the cognitive apparatus employed for language must have evolved for purposes other than speech production and comprehension, as it is just inconceivable that such a sophisticated cognitive apparatus could have evolved in as little as 200,000 years. The human language organ, in other words, must have evolved deep within the phylogenetic tree, suggesting that many of its features are still shared with non-human primates. Whether or not this scenario will withstand further testing remains to be seen. At present, it encourages a systematic investigation of the cognitive capacities of non-human primates during natural acts of communication to identify the phylogenetic roots of the cognitive capacities activated during language production and comprehension. Of particular interest are semantic and syntactic abilities, as they are central to virtually every definition of language (Crystal 1997).

Responses to semantic cues

Semantic competence in animals was first described in East African vervet monkeys (*Cercopithecus aethiops*). In this species, individuals produce acoustically distinctive alarm calls to at least five different types of predator: large terrestrial carnivores, eagles, snakes, baboons, and unfamiliar humans (Struhsaker 1967; Seyfarth & Cheney 1997). Playback experiments have demonstrated that some of these calls elicit anti-predator responses in other monkeys that resemble their natural responses to the corresponding predators (Seyfarth et al. 1980). Playbacks of eagle alarm calls, for instance, cause monkeys to look up into the air or run into a bush, while snake alarm calls may cause them to stand on their hind legs and scan the ground.

Meanwhile, comparable findings have also been reported from other primate species, notably from Diana and Campbell's monkeys (e.g. Ouattara et al. 2009, in press). Both species produce acoustically different alarm calls to crowned eagles and leopards. In both species, the alarm calls of the adult males and females are acoustically different from one another (Zuberbühler 2003). When hearing a Diana male's alarm calls nearby females respond with their own corresponding alarm calls that are acoustically very different, suggesting that the calls contain information about the type of predator present.

Like most forest monkeys, the Tai monkeys regularly forage in mixed species groups, presumably to improve their protection against predation (Noë & Bshary

1997). It is therefore not surprising that members of mixed species groups have also been observed to respond to each other's alarm calls, such as in the case of the Diana monkey – Campbell's monkey association (Wolters & Zuberbühler 2003). Playbacks of Diana monkey alarm calls to Campbell's monkeys and vice versa reliably elicited the appropriate alarm responses in the other species (Zuberbühler 2003; Figure 5).

The fact that a species is capable of producing different types of alarm calls is not a particularly exciting finding; it has been demonstrated in many groups of animals, primates and non-primates alike (e.g. *non-primates*: Slobodchikoff et al. 1991; Blumstein & Arnold 1995; Gyger et al. 1987; e.g. *primates*: Macedonia & Evans 1993; Fichtel & Kappeler 2002; Eckardt & Zuberbühler 2004). What is less well known is whether recipients distinguish the various types of alarm calls and, if so, whether they are able to link the different signals with particular events in the environment. These questions are non-trivial because different types of alarm

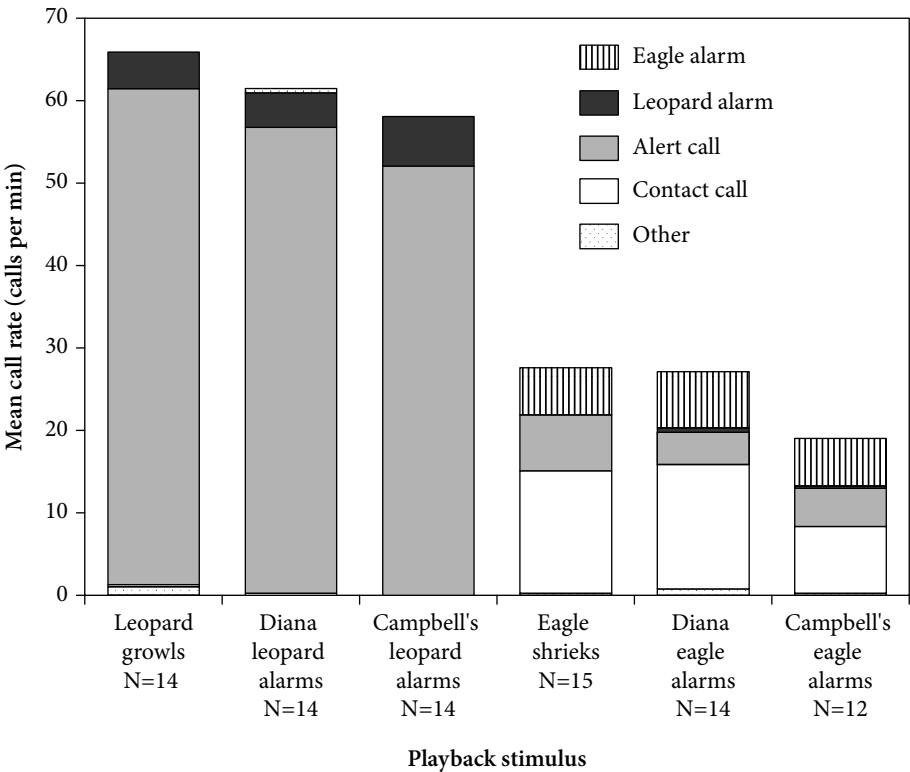


Figure 5. Cross-species communication between Diana and Campbell's monkeys in the Tai Forest, Ivory Coast. Diana monkeys respond the alarm calls of conspecifics in the same way as to alarm calls of Campbell's monkeys, provided the calls refer to the same predator

call may simply indicate whether or not a flight response is urgent, as in the case of California ground squirrels (*Spermophilus beecheyi*, Leger et al. 1980). One way of addressing this complicated issue is by carefully studying the recipients' responses to the various types of alarm calls. If recipients reliably respond in highly specific ways, then this suggests that they are processing the signals as indicators of external events. This clearly happens in Diana monkeys, but it is not necessarily a unique aptitude of primates. Playback experiments with yellow-casqued and black-casqued hornbills (*Ceratogymna elata* and *C. atrata*) have shown that these birds readily distinguish between monkey eagle and leopard alarm calls in a way that suggests that they attend to the calls' meanings (Rainey et al. 2004a, b; see Rasa 1983 for an additional example). How individuals acquire their knowledge about the relationship between an alarm call and a particular external event is an empirically little explored problem, although associative learning theory may be able to account for some of it (Rescorla & Wagner 1972).

Although the learning mechanisms are still poorly understood, the actual content of the monkeys' knowledge is somewhat better investigated, giving some insights into the mental representations that drive the monkeys' behaviour. In the case of the Diana monkeys, one question of particular interest was whether individuals, when hearing alarm calls, invoked a mental representation of the corresponding predator, or whether they were just responding to the acoustic features of the calls. Under field conditions, questions concerning mental processes are difficult to address, as the choice of experimental techniques is very restricted. One paradigm of significant value is the habituation-dishabituation procedure, initially developed for pre-linguistic children and later applied to free-ranging primates (Eimas et al. 1971; Cheney & Seyfarth 1988). Using one variation of this technique, several groups of Diana monkeys were first exposed to a prime stimulus, either a predator vocalization (baseline) or an alarm call (test and control). Five minutes later, the same group then heard the probe stimulus, the recording of a predator vocalisation, which was either semantically matching or non-matching to the prime stimulus. The prediction was that if the monkeys were able to invoke a mental representation of the predator while hearing an alarm call, then they should not be surprised to hear the semantically matching predator vocalisation used as a probe. Figure 6 illustrates the basic design and the main results.

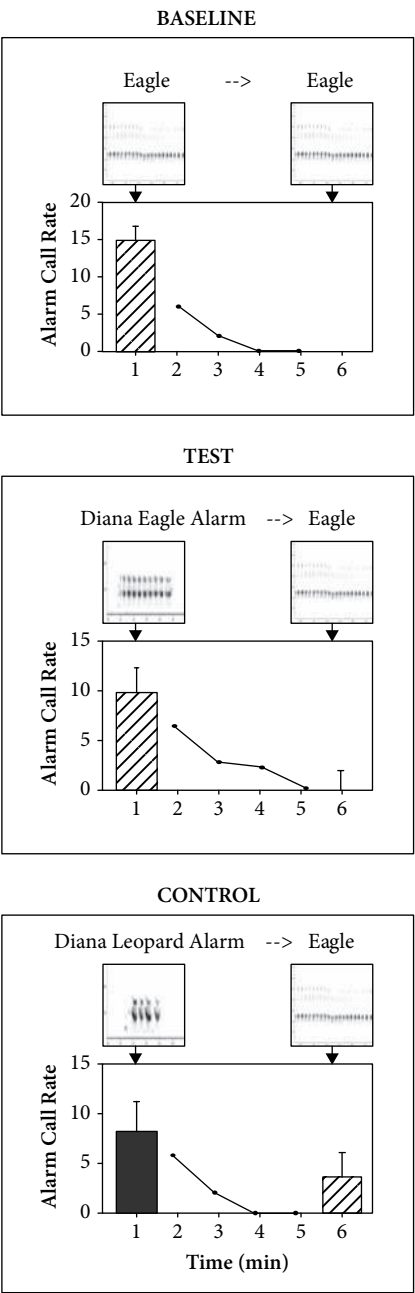


Figure 6. Design of a prime-probe experiment carried out with wild Diana monkeys to assess whether callers were able to recruit mental representations of the corresponding predator categories when producing the acoustically different alarm call types (Zuberbühler et al. 1999a)

Results showed that the semantic content of the prime stimuli, not their acoustic features alone, explained the response patterns of the monkeys to the probe stimuli (Zuberbühler et al. 1999a). That is, both eagle shrieks and leopard growls, two normally very powerful stimuli, lost their effectiveness in eliciting alarm calls when subjects were primed with the corresponding male alarm calls. Diana monkeys have also demonstrated their ability to process the semantic features of Campbell's alarm calls, analogous to how they process their own calls (Zuberbühler 2000b). It is currently not known whether non-primate species are capable of doing the same. In the case of the hornbills one would have to show that individuals do not directly respond to the monkeys' alarm calls, but that their behaviour is driven by a mental representation of the predator type indicated by the various alarm calls.

Responses to syntactic cues

Captive cotton-top tamarins are able to extract abstract rules from artificial meaningless stimuli presented in various patterns (Hauser et al. 2002, Fitch & Hauser 2004). In the field, when hearing "boom"-introduced Campbell's alarm calls, nearby Diana monkeys do not react by giving alarm calls themselves, which contrasts sharply with their vocal response to normal, that is "boom"-free, Campbell's alarm calls. These observations have led to the hypothesis that the booms may selectively affect the meaning of subsequent alarm calls. In one experiment, Campbell's eagle and Campbell's leopard alarm calls were combined with or without a pair of introductory booms. Although this is in itself an artificial combination, Diana monkeys responded significantly less strongly to boom-introduced alarm call series, suggesting that the booms have modified the semantic content of subsequent calls (Figure 7). This finding is particularly remarkable because Campbell's monkeys do not naturally combine boom calls with eagle or leopard alarm calls.

More recently, Arnold & Zuberbühler (2008) working with free-ranging putty-nosed monkeys in Nigeria played back different combinations of hacks and pyows from familiar and unfamiliar males to different females. Results showed that the females did not attend to individual calls when taking behavioural decisions, but that they attended to the entire call sequence presented to them, provided it came from their own male (Figure 8). This work thus demonstrates that, in primate communication, meaning can reside in a call, rather than individual calls.

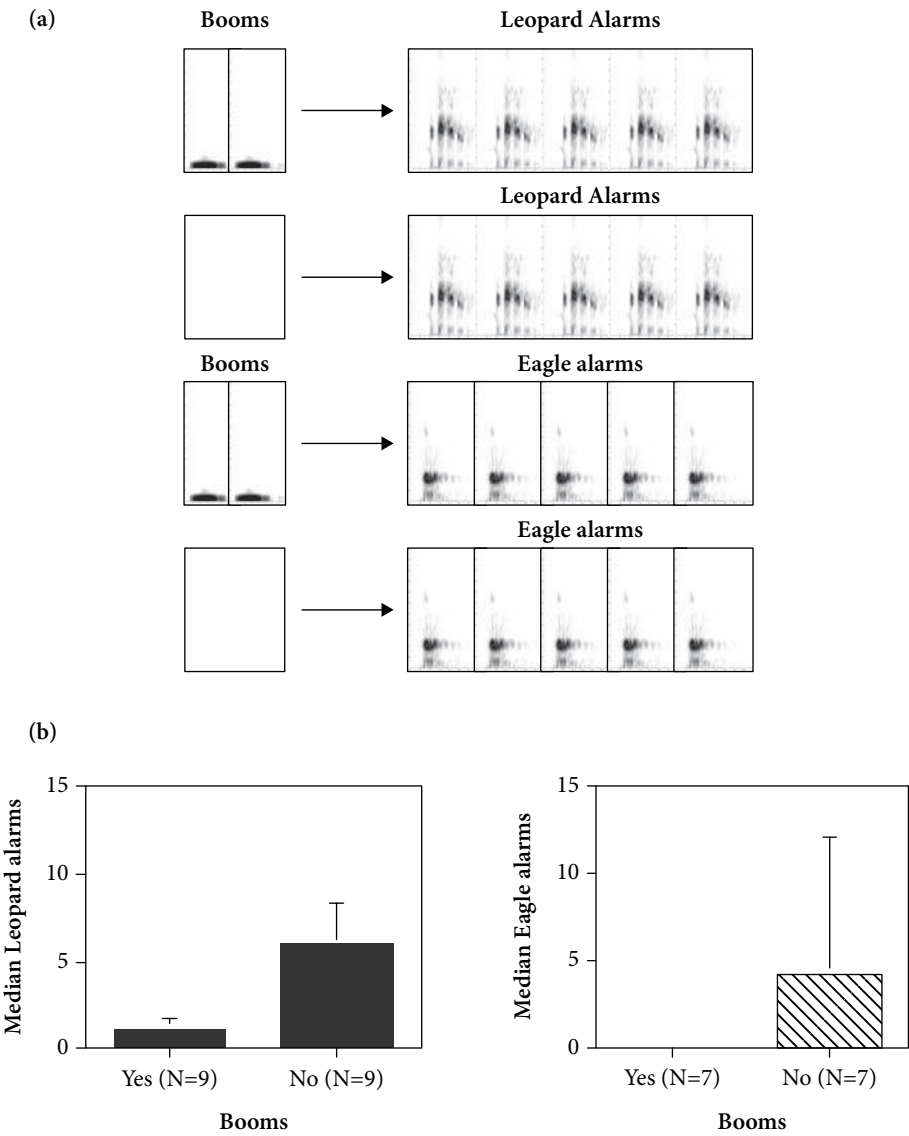


Figure 7. Responses of wild Diana monkeys to different alarm call sequences produced by Campbell’s monkeys. Adding “boom” calls in front of subsequent eagle or leopard alarm calls generally weakened the Diana monkeys’ response significantly (Zuberbühler 2002b)

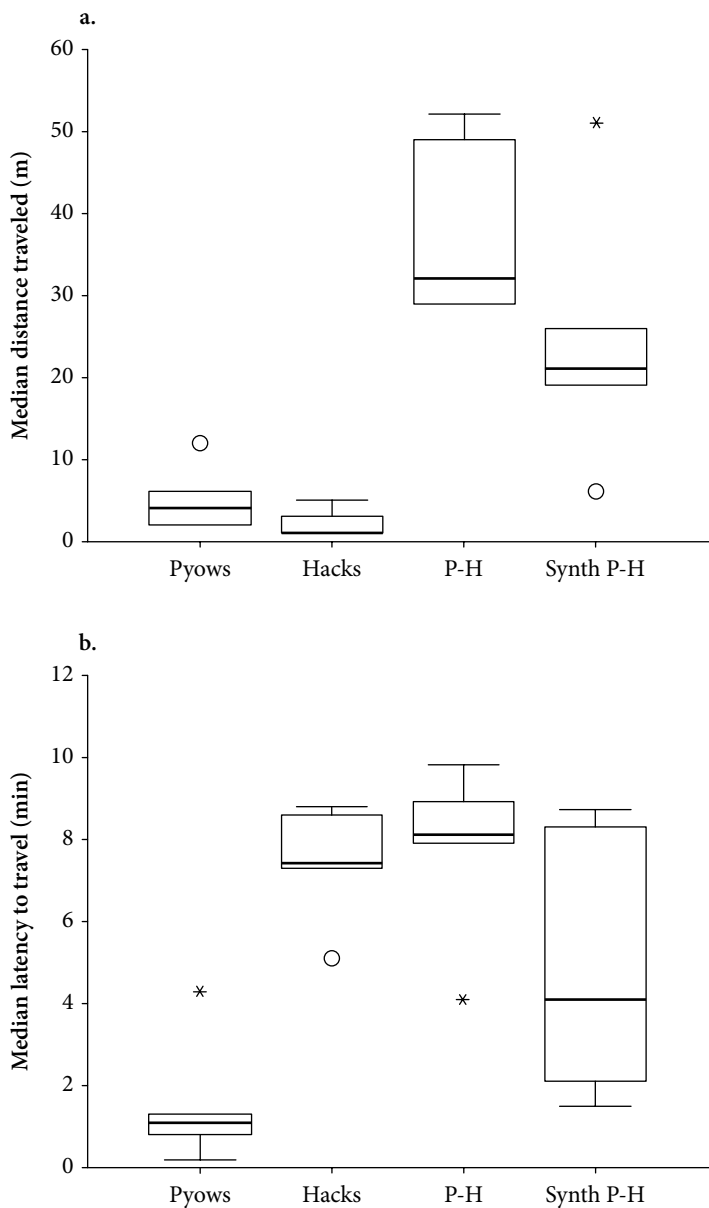


Figure 8. Responses of putty-nosed monkey females to different call sequences produced by their own males (a) travel distance, (b) response latency (Arnold & Zuberbühler 2008)

Responses to pragmatic cues

Many field researchers have reported that primates are highly attentive to the alarm calls of other species (Cheney & Seyfarth 1990). In the Taï forest, primates not only respond to the alarm calls of other primates, but also react to the vocalisations of several non-primate species (Zuberbühler & Jenny 2002). Crested guinea fowls (*Guttera pucherani*) are gregarious ground-dwelling forest birds that produce a rattling sounding alarm call during pursuit by a ground predator, such as leopards, which are fairly abundant in the Taï forest and which prey on birds (Zuberbühler & Jenny 2002). However, guinea fowls have also been observed to produce the same alarm calls in response to humans. For nearby monkeys these guinea fowl alarm calls thus provide a semantically ambiguous signal, because the most adaptive response to leopards (conspicuous alarm calling) is incompatible with the most adaptive response to humans (crypsis).

To investigate whether the monkeys take into account these complex relationships and to see whether they use their knowledge to take appropriate behavioural decisions, different groups of Diana monkeys were exposed to various stimulus pairs. A first set of groups was lead to believe that a leopard was present (by playing back brief recordings of leopard growls). A second set of groups experienced a brief recording of human speech to suggest the presence of poachers. After a 5-min period of silence, groups of both treatments heard a second playback, a series of guinea fowl alarm calls. If the monkeys were able to draw inferences about the potential causes of guinea fowl alarm calls, rather than just responding to the stimuli directly, then they should respond strongly to guinea fowl alarm calls only after being primed with leopard growls, but not after being primed with human speech (Figure 9).

There was a significant difference in the way leopard-primed and human-primed Diana monkey groups responded to guinea fowl alarm calls. Leopard-primed groups generally responded with a substantial number of alert calls to the guinea fowl stimulus, while human primed groups remained largely silent. No leopard alarm calls were given in either case, corroborating the findings of the previous prime-probe study (Zuberbühler et al. 1999a). In sum, results suggested that the monkeys' response was not directly driven by the guinea fowl alarm calls themselves, but by the monkeys' beliefs about the type of predator most likely to have caused the birds' alarm calls (Zuberbühler 2000c).

A similar problem exists when the monkeys are confronted with a nearby group of chimpanzees, another dangerous predator. Observations have shown that wild chimpanzees are occasionally attacked and preyed upon by leopards, although at lower rates compared to the monkeys (Zuberbühler & Jenny 2002). Chimpanzees give loud and conspicuous alarm calls in the presence of leopards, the 'SOS' screams, presumably to recruit help (Goodall 1986). Playbacks of chimpanzee

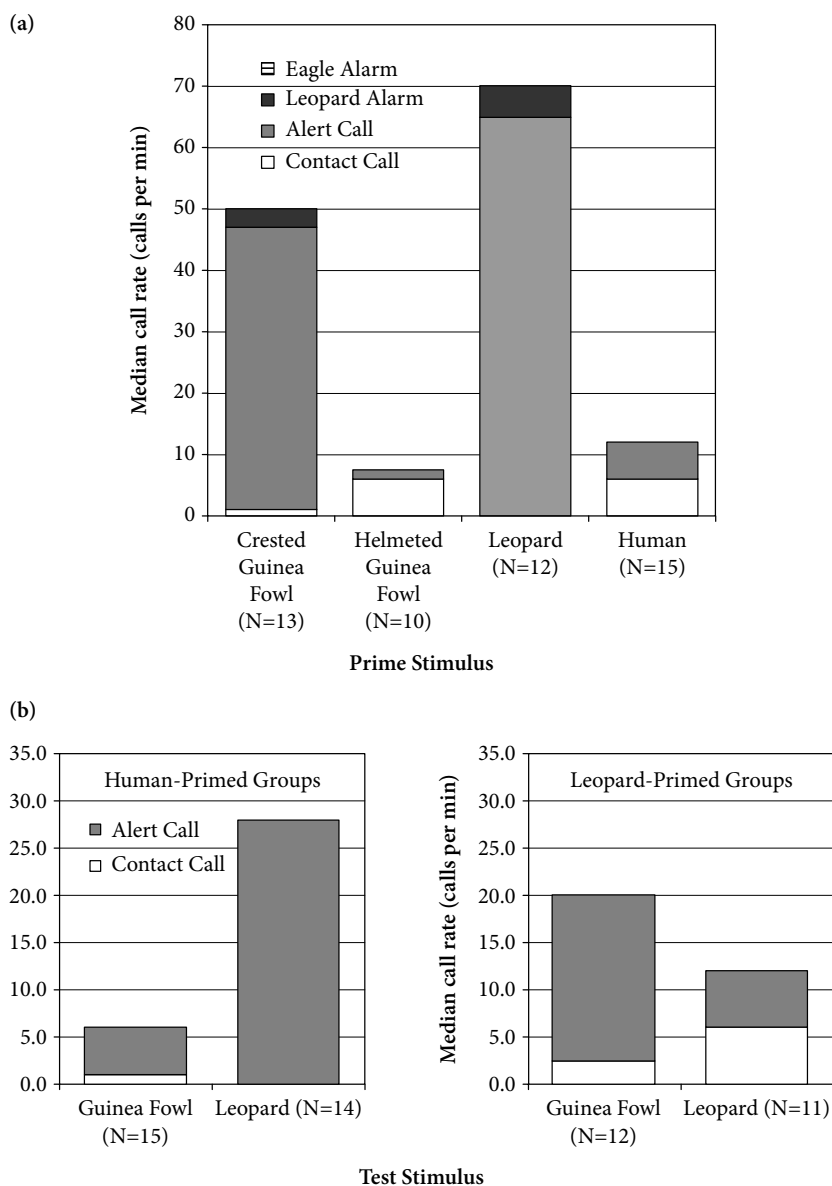


Figure 9. Top graph: Responses of wild Diana monkeys to crested and helmeted guinea fowl alarm calls compared to their vocal responses to two main predators, leopards and human poachers. Left bottom graph: If primed with human vocalisations Diana monkeys responded significantly weaker to guinea fowl alarm calls compared to leopard growls, as if they responded to the underlying cause of the calls, rather than the calls themselves. Right bottom graph: If primed with leopard growls, no differences were found in the monkeys' responses to guinea fowl alarm calls or leopard growls (Zuberbühler 2000c)

'SOS' screams to different groups of Diana monkeys revealed that about half of all the groups abandoned their typical chimp-specific cryptic response in favour of a leopard-specific conspicuous response, suggesting that some individuals assumed the presence of a leopard when hearing the chimpanzee alarm screams (Zuberbühler 2000d). Interestingly, Diana monkey groups with home ranges located in the core area of a resident chimpanzee community were significantly more likely to do so than peripheral groups. One interpretation of these findings is that the peripheral Diana monkey groups had less experience with chimpanzee vocalisations and, due to their less developed discrimination abilities, were more likely to respond cryptically to any type of chimpanzee vocalisation. It is clear that more rigorous testing will be required before any firm conclusions can be drawn. For instance, it is not clear whether the resident leopards use their range evenly and how much they avoid chimpanzee groups. Second, one might want to argue that the alarm call behaviour of adult males should not follow this pattern. This is due to the fact that, in this species, young males leave their native groups and settle elsewhere. In other words, males that settled with peripheral groups might have grown up in a core area group, suggesting that they should possess a superior knowledge of chimpanzee vocalisations. Data support this prediction. 8 of 19 peripheral males (42.1%) responded with leopard alarm calls when hearing chimpanzee screams, while this was only true for 5 of 19 groups of peripheral females (26.3%; Zuberbühler 2000d).

Vocal communication in apes

Surprisingly, comparably less work has been conducted on the natural communication of the apes (Marler & Tenaza 1977; Goodall 1986; Goodall 1968). This is especially puzzling given the various long-term field projects with chimpanzees, the phylogenetic proximity between humans and apes, and the apes' well documented skills in dealing with artificial communication systems (Premack 1970). If some monkey species exhibit a range of capacities that appear to have direct relevance to human language abilities, why is there no comparable evidence in the natural communication of our closest living relatives?

For these reasons we have started to study the vocal communication of free-ranging chimpanzees systematically (Slocombe & Zuberbühler *in press*). In one recent study we have identified a promising candidate for conveying social information in vocalisations used during conflicts, the screams. Wild chimpanzees at Budongo Forest, Uganda, produce screams and their acoustic structures reflect the two basic social roles that individuals can take during an agonistic interaction, the aggressor and the victim (Slocombe & Zuberbühler 2005a). For some time, it has been suspected that chimpanzee screams contain not only information about the

identity of the callers (Bauer & Philip 1983), but also social information, providing an important base for alliance decisions in other group members.

In addition, it has also emerged that the acoustic structure of victim screams reflects the severity of the attack experienced by the victim and that callers appear to have some control over this signal depending on the available audience. In the presence of high ranking group members, victims are more likely to produce screams that indicated a very severe aggression than in the absence of such an audience, regardless of the actual nature of the attack they had just experienced (Slocombe & Zuberbühler 2007). Naturalistic observations suggested that recipients used these acoustic differences conveyed by the screams as a basis for behavioural decisions, in particular whether or not to intervene in an ongoing conflict. We have recently been able to replicate these naturalistic observations with a field playback experiment (Slocombe et al. 2009). Work on chimpanzee copulation calls has further confirmed that chimpanzees are highly sensitive to the composition of their audience and gauge their vocal behaviour accordingly. For example, the females tend to refrain from producing copulation calls when other females are nearby, particularly if they are high-ranking (Townsend et al. 2008).

Another interesting type of vocal behaviour in chimpanzees concerns the response to finding food. Individuals then produce 'rough grunts' and behavioural manipulations have shown that the quality of the encountered food is reflected in the acoustic structure of these calls (Slocombe & Zuberbühler 2006). In a follow-up playback experiment, we were able to demonstrate that receivers were able to take advantage of these subtle acoustic differences and use them to make appropriate foraging decisions (Slocombe & Zuberbühler 2005b).

Gibbon songs, finally, have also been analysed in terms of semantic and syntactic properties. White-handed gibbons of Khao Yai National Park, Thailand, produce songs during normal morning duets but also to some ground predators, such as the clouded leopard. Predator model experiments have revealed that the composition of songs given to predators is subtly but consistently different from the songs given during normal morning duets, and some observations indicated that neighbouring groups were able to understand these differences in song composition (Clarke et al. 2006).

In sum, ongoing research with different ape species is beginning to reveal a number of patterns that suggest that vocal communication in these primates is comparable to what has already been found in different monkey species.

Conclusions

The studies reviewed in this chapter suggest that a number of properties of human language are also present in the natural communication systems of non-human primates. As signallers, primates are able to encode information about an external event and they can modify meaning of utterances based on structural rules and pragmatic information, including factors describing their audience. As receivers, they are very skilled in attending to different call types, sequences thereof, and they are able to integrate other information made available to them in order to extract meaning. Much of this evidence has been generated by work on monkeys (e.g. Cheney & Seyfarth 1990, 2007) but fieldwork on chimpanzees and other apes is also beginning to show comparable results.

Nevertheless, a number of fundamental discrepancies remain between human communication and that of the non-human primates. For example, several authors have argued that there may be a fundamental disconnection between the processes driving call production and those involved in call comprehension. Despite the various sources of flexibility identified earlier in this chapter, the main finding holds that, as signallers, primates appear to have little voluntary control over their articulators. They appear unable to create or imitate novel sound patterns by producing controlled articulatory movements during sustained larynx activity. Instead each species has evolved a relatively rigid set of call types. Although there can be acoustic variability within some of these call types, in most cases the different calls are only given in relatively specific contexts.

The underlying cognitive processes that lead to call production, however, do not necessarily have to be simple. In humans, laughter is also a fairly rigid behaviour over which the caller normally exerts relatively little voluntary control. Yet, whether or not an event is relevant enough to generate laughter is the result of sophisticated cognitive processes that draw on a range of mental representations and current assessment of the world and nearby audience.

Equally important, non-human primates generally appear to show little sensitivity to the psychological states of others, and much of this is reflected in their vocal behaviour. Some potentially interesting exceptions have been discussed in this chapter. For this reason, it has been argued that, in contrast to 'true' or intentional referentiality, which presumably underlies human speech, referential behaviour in animals merely functions in this way, while the underlying psychological processes might be different (Tomasello et al. 2005). It has been argued that humans possess language, not because they have specially evolved neural hardware that deals with linguistic problems, but due to their highly developed social abilities (Tomasello 2008). Humans are clearly unique in the degree to which they perceive each other's psychological states, a characteristic that may be the main

driving force to develop complex communication. According to this view, spoken language is just another cultural artefact of human existence, but one that builds on a number of biological predispositions humans brought along as part of their primate heritage.

How exactly humans have evolved their special adeptness for social behaviour is another unresolved problem. It is noteworthy that humans are the only primate species in which individuals regularly and on a large scale engage in collective actions with multiple unrelated individuals, opening avenues for the emergence of high levels of social behaviour amongst unrelated individuals (Smith 2003, Richardson et al. 2003).

Acknowledgments

We are grateful to the European Science Foundation's EUROCORES programme "OMLL", which allowed us to run the Tai Monkey Project in the Ivory Coast where much of this research has been conducted. Other significant funding has come from the BBSRC, the Leverhulme Trust, the EC FP6 initiative "What it means to be human", the British Academy, and the British Council.

References

- Arnold, K., Pohlner, Y. & Zuberbühler, K. 2008. A forest monkey's alarm call series to predator models. *Behavioral Ecology and Sociobiology*, **62**, 549–559.
- Arnold, K., Pohlner, Y. & Zuberbühler, K. 2009. The psychology of vocal behaviour in a forest primate. In: *The monkeys and apes of Gashaka: Primate socio-ecology and conservation in Nigeria's biodiversity hotspot* (Ed. by Sommer, V. & Ross, C.). New York: Springer.
- Arnold, K. & Zuberbühler, K. 2006a. The alarm-calling system of adult male putty-nosed monkeys, *Cercopithecus nictitans martini*. *Animal Behaviour*, **72**, 643–653.
- Arnold, K. & Zuberbühler, K. 2006b. Semantic combinations in primate calls. *Nature*, **441**, 303–303.
- Arnold, K. & Zuberbühler, K. 2008. Meaningful call combinations in a non-human primate. *Current Biology*, **18**, R202–R203.
- Bauer, H. R. & Philip, M. M. 1983. Facial and vocal individual recognition in the common chimpanzee. *Psych. Rec.*, **33**, 161–170.
- Blumstein, D. T. & Arnold, W. 1995. Situational specificity in alpine-marmot alarm communication. *Ethology*, **100**, 1–13.
- Brown, E. D. & Farabaugh, S. M. 1997. What birds with complex social relationships can tell us about vocal learning: Vocal sharing in avian groups. *Social influences on vocal development*, 98–127.
- Cheney, D. L. & Seyfarth, R. M. 1988. Assessment of meaning and the detection of unreliable signals by vervet monkeys. *Animal Behaviour*, **36**, 477–486.

- Cheney, D. L. & Seyfarth, R. M. 1990. *How monkeys see the world: Inside the mind of another species*. Chicago: Chicago University Press.
- Cheney, D. L. & Seyfarth, R. M. 2007. *Baboon metaphysics*. Chicago: University of Chicago Press.
- Clarke, E., Reichard, U. & Zuberbühler, K. 2006. The syntax and meaning of wild gibbon songs. *PLoS One*, **1**, e73.
- Cords, M. 1988. Mating systems of forest guenons: a preliminary review. In: *A primate radiation: Evolutionary biology of the African guenons* (Ed. by Gautier-Hion, A., Bourliere, F., Gautier, J.-P. & Kingdon, J.), pp. 323–. Cambridge: CUP.
- Crystal, D. 1997. *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.
- Eckardt, W. & Zuberbühler, K. 2004. Cooperation and competition in two forest monkeys. *Behavioral Ecology*, **15**, 400–411.
- Eimas, P. D., Siqueland, P., Jusczyk, P. & Vigorito, J. 1971. Speech perception in infants. *Science*, **212**, 303–306.
- Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., Monaco, A. P. & Pääbo, S. 2002. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, **418**, 869 – 872.
- Fant, G. 1960. *Acoustic theory of speech production*. The Hague: Mouton.
- Fichtel, C. & Kappeler, P. M. 2002. Anti-predator behavior of group-living Malagasy primates: Mixed evidence for a referential alarm call system. *Behavioral Ecology and Sociobiology*, **51**, 262–275.
- Fitch, W. T. 1997. Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America*, **102**, 1213–1222.
- Fitch, W. T. 2000. The phonetic potential of nonhuman vocal tracts: Comparative cineradiographic observations of vocalizing animals. *Phonetica*, **57**, 205–218.
- Fitch, W. T. & Hauser, M. D. 2004. Computational constraints on syntactic processing in a non-human primate. *Science*, **303**, 377–380.
- Geissmann, T. 1984. Inheritance of song parameters in the gibbon song, analysed in 2 hybrid gibbons (*Hylobates pileatus* x *H. lar*). *Folia primatologica*, **24**, 216–235.
- Ghazanfar, A. A. & Hauser, M. D. 1999. The neuroethology of primate vocal communication: substrates for the evolution of speech. *Trends in cognitive science*, **3**, 377–384.
- Goodall, J. 1968. A preliminary report on expressive movements and communication in the Gombe Stream chimpanzees. In: *Primate studies in adaptation and variability* (Ed. by Jay, P.). New York: Holt, Rinehart and Winston.
- Goodall, J. 1986. *The chimpanzees of Gombe: patterns of behavior*. Cambridge: Harvard University Press.
- Gyger, M., Marler, P. & Pickert, R. 1987. Semantics of an avian alarm call system: The male domestic fowl, *G. domesticus*. *Behaviour*, **102**, 15–40.
- Hailman, J. P. & Ficken, M. S. 1987. Combinatorial animal communication with computable syntax: Chick-a-dee calling qualifies as 'language' by structural linguistics. *Animal Behaviour*, **34**, 1899–1901.
- Hammerschmidt, K., Freudenstein, T. & Jürgens, U. 2001. Vocal development in squirrel monkeys. *Behaviour*, **138**, 1179–1204.
- Hausberger, M., Richard-Yris, M. A., Henry, L., Lepage, L. & Schmidt, I. 1995. Song sharing reflects the social organization in a captive group of European starlings (*Sturnus vulgaris*). *Journal of Comparative Psychology*, **109**, 222–241.
- Hauser, M. D., Evans, C. S. & Marler, P. 1993. The role of articulation in the production of rhesus monkey, *Macaca mulatta*, vocalizations. *Animal Behaviour*, **45**, 423–433.

- Hauser, M. D. & Schön-Ybarra, M. S. 1994. The role of lip configuration in monkey vocalizations – Experiments using Xylocaine as a nerve block. *Brain and Language*, **46**, 232–244.
- Hauser, M. D., Weiss, D. & Marcus, G. 2002. Rule learning by cotton-top tamarins. *Cognition*, **86**, B15–B22.
- Janik, V. M. 2000. Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*). *Science*, **289**, 1355–1357.
- Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, **413**, 519–523.
- Leger, D. W., Owings, D. H. & Gelfand, D. L. 1980. Single note vocalizations of California ground squirrels: graded signals and situation – specificity of predator and socially evoked calls. *Zeitschrift für Tierpsychologie*, **52**, 227–246.
- Lemasson, A. 2003. Communication vocale et organisation sociale chez la mone de Campbell (*Cercopithecus campbelli*). Partage vocal et relations sociales. In: *UMR CNRS 6552*. Rennes: Université de Rennes.
- Lemasson, A., Blois-Heulin, C., R., J. & Hausberger, M. 2006. Female social relationships in a captive group of Campbell's monkeys. *American Journal of Primatology*, **68**, 1161–1170.
- Lemasson, A., Gautier, J. P. & Hausberger, M. 2003. Vocal similarities and social bonds in Campbell's monkey (*Cercopithecus campbelli*). *Comptes Rendus Biologies*, **326**, 1185–1193.
- Lemasson, A. & Hausberger, M. 2004. Patterns of vocal sharing and social dynamics in a captive group of Campbell's monkeys (*Cercopithecus campbelli*). *J. Comp. Psych.*
- Lemasson, A., Hausberger, M. & Zuberbühler, K. 2005. Socially meaningful vocal plasticity in adult Campbell's monkeys (*Cercopithecus campbelli*). *Journal of Comparative Psychology*, **119**, 220–229.
- Lemasson, A., Richard, J. P. & Hausberger, M. 2004. A new methodological approach to context analysis of call production. *Bioacoustics*.
- Lieberman, D. E., McCarthy, R. C., Hiimeae, K., Lieberman, P. & Palmer, J. B. 1998. New estimates of fossil hominid vocal tract dimensions. *Journal of Human Evolution*, **34**, A12–A13.
- Lieberman, P. 1969. On the acoustic analysis of primate vocalizations. *Behavior Research Methods & Instrumentation*, **1**, 169–174.
- Lieberman, P., Klatt, D. H. & Wilson, W. H. 1969. Vocal tract limitations on the vowel repertoires of rhesus monkeys and other nonhuman primates. *Science*, **164**, 1185–1187.
- Macedonia, J. M. & Evans, C. S. 1993. Variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethology*, **93**, 177–197.
- Marler, P. 1970. Birdsong and speech development: could there be parallels? *Am. Sci.*, **58**, 669–673.
- Marler, P. & Tenaza, R. 1977. Signaling behavior of apes with special reference to vocalizations. In: *How animals communicate* (Ed. by Sebeok, T. A.), pp. 965–1033 (as copies: 965, 971, 982, 983, 986–989). Bloomington: Indiana Univ. Press.
- Marshall, A., Wrangham, R. & Clark Arcadi, A. 1999. Does learning affect the structure of vocalizations in chimpanzees? *Animal Behaviour*, **58**, 825–830.
- McGraw, S. W., Zuberbühler, K. & Noë, R. 2007. Monkeys of the Taï Forest: An African Monkey Community. In: *Cambridge Studies in Biological and Evolutionary Anthropology* (No. 51). Cambridge: Cambridge Univ. Press.
- Milroy, L. 1987. *Language and social networks*. Oxford: Basil Blackwell.
- Negus, V. E. 1949. *The comparative anatomy and physiology of the larynx*. New York: Hafner.
- Noë, R. & Bshary, R. 1997. The formation of red colobus-diana monkey associations under predation pressure from chimpanzees. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **264**, 253–259.

- Nowicki, S. 1987. Vocal tract resonances in oscine bird sound production: Evidence from bird-songs in a helium atmosphere. *Nature*, **325**, 53–55.
- Ouattara, K., Lemasson, A. & Zuberbühler, K. in press. Anti-predator strategies of free-ranging Campbell's monkeys. *Behaviour*.
- Ouattara, K., Zuberbühler, K., N'goran, E. K., Gobert, J.-E. & Lemasson, A. 2009. The alarm call system of female Campbell's monkeys. *Animal Behaviour*, **78**, 35–44.
- Owren, M. J. & Bernacki, R. H. 1998. Applying linear predictive coding (LPC) to frequency-spectrum analysis of animal acoustic signals. In: *Animals Acoustic Communication* (Ed. by Hopp, S. L., Owren, M. J. & Evans, C. S.), pp. 129–162. Berlin: Springer.
- Owren, M. J., Dieter, J. A., Seyfarth, R. M. & Cheney, D. L. 1993. Vocalizations of rhesus (*Macaca mulatta*) and japanese (*M. fuscata*) macaques cross-fostered between species show evidence of only limited modification. *Developmental Psychobiology*, **26**, 389–406.
- Owren, M. J., Seyfarth, R. M. & Cheney, D. L. 1997. The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cynocephalus ursinus*): Implications for production processes. *Journal of the Acoustical Society of America*, **101**, 2951–2963.
- Papworth, S., Böse, A. S., Jessica Barker, J. & Zuberbühler, K. 2008. Male blue monkeys alarm call in response to danger experienced by others. *Biology Letters*, in press.
- Premack, D. 1970. A functional analysis of language. *J. Exp. Analysis Behav.*, **14**, 107–125.
- Rainey, H. J., Zuberbühler, K. & Slater, P. J. B. 2004b. The responses of black-casqued hornbills to predator vocalisations and primate alarm calls. *Behaviour*, **141**, 1263–1277.
- Rainey, H. J., Zuberbühler, K. & Slater, P. J. B. 2004a. Hornbills can distinguish between primate alarm calls. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **271**, 755–759.
- Range, F. & Fischer, J. 2004. Vocal repertoire of sooty mangabeys (*Cercocebus torquatus atys*) in the Tai National Park. *Ethology*, **110**, 301–321.
- Rasa, O. A. E. 1983. Dwarf mongoose and hornbill mutualism in the Taru desert, Kenya. *Behav. Ecol. Sociobiol.*, **12**, 181–190.
- Rescorla, R. A. & Wagner, A. R. 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: *Classical conditioning II: Current research and theory* (Ed. by Black, A. & Prokasy, W.), pp. 64–69. New York: Appleton-Century-Crofts.
- Richerson, P. J., Boyd, R. & Henrich, J. 2003. Cultural evolution of human cooperation. In: *Genetic and cultural evolution of cooperation* (Ed. by Hammerstein, P.), pp. 357–388. Cambridge, MA: MIT Press.
- Riede, T., Bronson, E., Hatzikirou, H. & Zuberbühler, K. 2005. Vocal production mechanisms in a non-human primate: morphological data and a model. *Journal of Human Evolution*, **48**, 85–96.
- Riede, T., Bronson, E., Hatzikirou, H. & Zuberbühler, K. 2006. Multiple discontinuities in non-human vocal tracts – A response to Lieberman (2006). *Journal of Human Evolution*, **50**, 222–225.
- Riede, T. & Fitch, T. 1999. Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*). *J. Exp. Biol.*, **202**, 2859–2867.
- Riede, T. & Zuberbühler, K. 2003a. The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization. *Journal of the Acoustical Society of America*, **114**, 1132–1142.
- Riede, T. & Zuberbühler, K. 2003b. Pulse register phonation in Diana monkey alarm calls. *J. Acoust. Soc. Am.*, **113**, 2919–2926.

- Roush, R. S. & Snowdon, C. T. 1999. The effects of social status on food-associated calling behaviour in captive cotton-top tamarins. *Animal Behaviour*, **58**, 1299–1305.
- Schel, A. M., Tranquilli, S. & Zuberbühler, K. 2009. The alarm call system of black-and-white colobus monkeys. *J Comp Psychol*.
- Seyfarth, R. M. & Cheney, D. L. 1997. Behavioral mechanisms underlying vocal communication in nonhuman primates. *Animal Learning & Behavior*, **25**, 249–267.
- Seyfarth, R. M., Cheney, D. L. & Marler, P. 1980. Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, **210**, 801–803.
- Slobodchikoff, C. N., Kiriazis, J., Fischer, C. & Creef, E. 1991. Semantic information distinguishing individual predators in the alarm calls of Gunnison's prairie dogs. *Animal Behaviour*, **42**, 713–719.
- Slocombe, K. E., Townsend, S. W. & Zuberbühler, K. 2009. Wild chimpanzees (*Pan troglodytes schweinfurthii*) distinguish between different scream types: evidence from a playback study. *Animal Cognition*, **12**, 441–449.
- Slocombe, K. E. & Zuberbühler, K. 2005a. Agonistic screams in wild chimpanzees (*Pan troglodytes schweinfurthii*) vary as a function of social role. *Journal of Comparative Psychology*, **119**, 67–77.
- Slocombe, K. E. & Zuberbühler, K. 2005b. Functionally referential communication in a chimpanzee. *Current Biology*, **15**, 1779–1784.
- Slocombe, K. E. & Zuberbühler, K. 2006. Food-associated calls in chimpanzees: responses to food types or food preferences? *Animal Behaviour*, **72**, 989–999.
- Slocombe, K. E. & Zuberbühler, K. 2007. Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 17228–17233.
- Slocombe, K. E. & Zuberbühler, K. in press. Vocal communication in chimpanzees. In: *The Mind of the Chimpanzee: Ecological and Empirical Perspectives* (Ed. by Lonsdorf, E. V., Ross, S. & Matsuzawa, T.). Chicago: University of Chicago Press.
- Smith, E. A. 2003. Human cooperation. Perspectives from behavioral ecology. In: *Genetic and cultural evolution of cooperation* (Ed. by Hammerstein, P.), pp. 83–93. Cambridge MA: MIT Press.
- Snowdon, C. T. & Elowson, A. M. 1999. Pygmy marmosets modify call structure when paired. *Ethology*, **105**, 893–908.
- Stephan, C. & Zuberbühler, K. 2008. Predation increases acoustic complexity in primate alarm calls. *Biology Letters*, **4**, 641–644.
- Struhsaker, T. T. 1967. Auditory communication among vervet monkeys (*Cercopithecus aethiops*). In: *Social communication among primates* (Ed. by Altmann, S. A.), pp. 281–324. Chicago: University of Chicago Press.
- Sutton, D., Larson, C., Taylor, E. M. & Lindeman, R. C. 1973. Vocalization in rhesus monkey: Conditionability. *Brain Research*, **52**, 225–231.
- Tomasello, M. 2008. *Origins of human communication*. Cambridge MA: MIT Press.
- Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. 2005. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, **28**, 675–+.
- Tomasello, M. & Zuberbühler, K. 2002. Primate vocal and gestural communication. In: *The cognitive animal* (Ed. by Bekoff, M., Allen, C. & Burghardt, G. M.), pp. 293–299. Cambridge MA: MIT Press.
- Townsend, S. W., Deschner, T. & Zuberbühler, K. 2008. Female chimpanzees use copulation calls flexibly to prevent social competition. *PLoS One*, **3**, e2431.

- Townsend, S. W., Slocombe, K. E., Thompson, M. E. & Zuberbühler, K. 2007. Female-led infanticide in wild chimpanzees. *Current Biology*, **17**, R355–R356.
- Townsend, S. W. & Zuberbühler, K. in press. Audience effects in chimpanzee copulation calls. *Communicative & Integrative Biology*.
- von Humboldt, K. W. 1836/1935. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Berlin: Druckerei der Königlichen Akademie der Wissenschaften.
- Wich, S. A. & de Vries, H. 2006. Male monkeys remember which group members have given alarm calls. *Proceedings of the Royal Society B-Biological Sciences*, **273**, 735–740.
- Winter, P., Handley, P., Ploog, D. & Schott, D. 1973. Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behaviour*, **47**, 230–239.
- Wolters, S. & Zuberbühler, K. 2003. Mixed-species associations of Diana and Campbell's monkeys: The costs and benefits of a forest phenomenon. *Behaviour*, **140**, 371–385.
- Zuberbühler, K. 2000a. Referential labelling in Diana monkeys. *Animal Behaviour*, **59**, 917–927.
- Zuberbühler, K. 2000b. Interspecific semantic communication in two forest monkeys. *Proc. R. Soc. Lond. B*, **267**, 713–718.
- Zuberbühler, K. 2000c. Causal cognition in a non-human primate: field playback experiments with Diana monkeys. *Cognition*, **76**, 195–207.
- Zuberbühler, K. 2000d. Causal knowledge of predators' behaviour in wild Diana monkeys. *Animal Behaviour*, **59**, 209–220.
- Zuberbühler, K. 2002a. Effects of natural and sexual selection on the evolution of guenon loud calls. In: *The guenons: diversity and adaptation in African monkeys* (Ed. by Glenn, M. E. & Cords, M.), pp. 289–306. New York: Plenum.
- Zuberbühler, K. 2002b. A syntactic rule in forest monkey communication. *Animal Behaviour*, **63**, 293–299.
- Zuberbühler, K. 2003. Referential signalling in non-human primates: cognitive precursors and limitations for the evolution of language. *Advances in the study of behavior*, **33**, 265–307.
- Zuberbühler, K. 2005. The phylogenetic roots of language – Evidence from primate communication and cognition. *Current Directions in Psychological Science*, **14**, 126–130.
- Zuberbühler, K. 2008a. Audience effects. *Current Biology*, **18**, RXXX.
- Zuberbühler, K. 2008b. Referentiality and concepts in animal cognition. In: *Encyclopedia of Neuroscience* (Ed. by Squire, L. R.). Oxford: Academic Press.
- Zuberbühler, K., Cheney, D. L. & Seyfarth, R. M. 1999a. Conceptual semantics in a nonhuman primate. *Journal of Comparative Psychology*, **113**, 33–42.
- Zuberbühler, K. & Jenny, D. 2002. Leopard predation and primate evolution. *Journal of Human Evolution*, **43**, 873–886.
- Zuberbühler, K. & Jenny, D. 2007. Interactions between leopards and monkeys. In: *Monkeys of the Tai Forest: An African monkey community* (Ed. by McGraw, S., Zuberbühler, K. & Noë, R.). Cambridge: Cambridge Univ. Press.
- Zuberbühler, K., Jenny, D. & Bshary, R. 1999b. The predator deterrence function of primate alarm calls. *Ethology*, **105**, 477–490.
- Zuberbühler, K., Noë, R. & Seyfarth, R. M. 1997. Diana monkey long-distance calls: messages for conspecifics and predators. *Animal Behaviour*, **53**, 589–604.

Language evolution and computer modeling

Can agent-based language evolution contribute to archeology?

Luc Steels¹

A growing scala of computational and robotic experiments are trying to pin down the cognitive and social prerequisites that may have given rise to human language. From humble beginnings showing how a lexicon may self-organize in a population of artificial agents, these research efforts are now exploring how grammatical languages about complex scenes may emerge. This paper introduces this field of inquiry and then explores whether a dialog with archeologists might be useful.

Introduction

Modeling is an essential tool in all sciences and it has also a contribution to make to the study of the origins and evolution of human languages. Modeling can help us understand what kind of cognitive functions and interaction patterns are necessary and sufficient for the origins and evolution of language and examine through mathematical investigations and computational simulations whether certain basic assumptions of a theory are viable or not.

Making models of language origins and evolution seems at first sight very difficult, if not impossible. However, during the past decade, significant progress has been made by a growing group of researchers producing computational simulations and robotic experiments in which artificial agents bootstrap unaided and *from scratch* communication systems with similarities to human natural languages. These modeling efforts help to test very concrete hypotheses of the cognitive and social prerequisites that are required to bootstrap a language system and preserve it through cultural evolution. Why would this be of any interest to archeologists and anthropologists?

There is no direct empirical evidence of the very beginning of language and there will never be, but by comparing the cognitive and social prerequisites suggested

1. Sony Computer Science Laboratory – Paris, Vrije Universiteit Brussel

by modeling work with the cognitive capacities of early humans as suggested by artefacts and evidence of population dynamics and migration patterns, it is possible to compare and thus validate or question hypotheses independently arrived at from both directions, just as evidence from genetics can be paired with archeological evidence from migration patterns to strengthen hypotheses derived from each side. This is a very common and highly desirable situation in science. For example, climate reconstructions based on fluid dynamics models are compared with data derived independently from ice cores, tree rings, and other indices of climate change, in order to help validate hypotheses and conclusions derived by both methods.

The next section of the paper provides first an overview of the different types of models that have been developed. We then focus on one class of models, those assuming socio-cultural as opposed to genetic origins, give a broad idea of the current scope of these models and provide some pointers to the literature. The remainder of the paper then formulates some of the hypotheses underlying the models in such a way as to start the dialog with archeologists.

Agent-based models

Modeling efforts can be classified along two dimensions: The first dimension distinguishes *aggregate* versus *agent-based* models.

1. Aggregate models do not model the internals of every individual language user explicitly but aggregate the behavior of a population and then formulate equations defining a time evolution among these aggregate quantities (see for example Abrams(2003)). This kind of approach is particularly useful to study the effect of population dynamics or population network structure on language dynamics.
2. *Agent-based models* explicitly model all the internal linguistic structures and cognitive processes of 'language-ready' agents as well as the grounded interaction patterns of these agents, in order to show what the necessary and sufficient conditions are to see the formation of symbolic communication systems and their preservation across generations (see examples in MinettWang(2005), Lyon(2007)).

To have enough realism, agent-based models have to be rather complex, and hence they require computer simulations to examine their consequences. Moreover to study the effect of embodiment, grounding, and real world interaction on an emergent language system, experiments with robots that bootstrap their own language systems through situated embodied interactions are highly instructive, if not

necessary, and this is also what some researchers, particularly in my own group, have been doing Steels(2005).

Using a modeling approach to investigate language evolution does not imply any particular theoretical stance. Any theory could (indeed should) be amenable to explicit modeling as long as it is sufficiently complete and coherent. However modeling approaches have already shown that some theoretical assumptions work better than others and that some quite common assumptions ignore critical aspects of the problem.

There is a wide consensus that language rests on three aspects: biology, cognition, and culture. Biology concerns the neuronal and physiological hardware, cognition concerns the information representation, processing and learning supported by this hardware, and culture concerns the shared system that emerges out of the linguistic activities of individuals as it persists over time. On the other hand, there is no consensus on which one of these aspects is the dominant force. Consequently we get a second dimension for classifying models, depending on which aspect they take as primary.

1. *Biology*: Some researchers have emphasized the role of genetic evolution through natural selection in the origins of language (see for example Pinker and Jackendoff(2005), Bickerton(1984)). They hypothesize that the language faculty is a highly specialized, genetically coded organ in the brain, which evolved through genetic evolution by natural selection. In this view language is acquired mainly by setting parameters in the innate schemata provided by Universal Grammar (Thornton and Wexler,1999), rather than through inductive learning or problem solving. Although this nativist approach to language origins has been explored in a variety of experiments using the frameworks of genetic algorithms (Cangelosi and Parisi(1998), Szathmari(2007)) it has turned out to be very difficult to get convincing simulations off the ground for explaining the origins of the grammatical features of languages.

2. *Culture*: Some researchers propose that cultural evolution is the primary force. They hypothesize that human neurobiology provides both general learning mechanisms *and* language-specific biases that speed up learning in order to overcome the 'poverty of the stimulus' bottleneck Elman(1996), Kirby and Hurford(2002). Recent computational and mathematical models have indeed shown that if inductive learning is chained through a sequence of vertical transmissions from teacher to learner, language structures and conceptualizations gradually appear that reflect the inductive bias of the language learners (Griffiths,2005). So this route is promising. However, the problem with these models is that the desired grammatical structure has already to be present in the learning bias, which raises the question where the original bias may have come from. Moreover these models do not

address how coherence may arise and be maintained in the population despite unavoidable language variation, or how a language may keep adapting itself to the changing needs of its users Hopper(1987).

3. *Cognition*: Others put the main emphasis on social interaction patterns and cognition (see examples in Heine (1997), Tomasello(1999), Batali(1998), Steels(2005)). They argue that human neurobiology supports a large battery of cognitive functions usable in a wide range of domains, and that language has recruited and then exploits many of these cognitive functions to the fullest. Language production and interpretation are hence viewed as problem solving processes, alternating between routine problem solving and the creative invention of new conceptualizations or language forms in order to handle novel situations or make production or interpretation become more efficient. These innovations become entrenched through well-documented grammaticalization processes (Heine and Kuteva,2008). In this view, language learning is viewed as a problem solving process that engages all cognitive resources of the learner, including the ability to form new categories, to guess meaning through inference, to apply analogies and metaphors, to hypothesize the function of novel syntactic structures, to generate and test out new phonetic features, etc. There is still a strong cultural component in these models because individuals pick up language constructs from each other and features may get lost in inter-generational learning, but this is no longer the sole force in generating linguistic structure.

Socio-cognitive models are necessarily quite complex as they require modeling and operationalizing many aspects of cognition, including perception, memory, conceptualization, planning and plan recognition, metaphor, analogy, joint attention, perspective reversal, etc. But without adequate precise information processing models of the processes that go into language production, language comprehension and language learning we cannot know whether these processes are unique to language or shared by a wide spectrum of cognitive tasks. We cannot know what can be learned and what needs to be provided innately as 'principles and parameters' or as innate biases in intergenerational cultural evolution. Of course, in the end the three forces (biology, cognition, and culture) play a role and they interact and impinge upon each other, but for the purpose of scientifically advancing the field, it is useful to focus on one force and see what kind of explanations it can generate.

The work in my laboratory over the past decade has focused on developing agent-based socio-cultural models of language origins using computational simulations and robotic experiments (Steels, 2003) and this will also be the point of view underlying the rest of the paper. Our main thesis is that language is a (cultural) invention based on similar cognitive capacities as found in other domains where humans excel or are special, such as tool use, collective hunting, social organisation (kinship structures, social hierarchy, societal structure), or spatial navigation.

Besides the proven experimental successes, there are three reasons why the socio-cultural hypothesis appears more plausible than a purely genetic origins driven by natural selection.

1. If we just take spatial navigation for example, we clearly see that similar problem solving and learning skills are needed compared to language: In order to perform the spatial and geographic reasoning required for navigation, individuals must become sensitive to the significant features of the landmarks in their environment, a sequence of actions must be planned and their execution monitored, actions to navigate from one place to another often have a hierarchical structure as more complex paths are conceived by combining subpaths, after finding one subpath it may be stored for later use so that aspects of navigation becomes more routine and more challenging journeys can be undertaken. All of these capacities are needed for language as well. There is even a social dimension to navigation: Individuals are able to recognise the paths being executed by others or complete and correct someone's partial path once begun. They can explain to another person how to get somewhere, even with gestures or drawings. Spatial navigation is therefore one example of a cognitive domain with a lot of similarities to language, because it is also a skill that develops progressively and requires constant adaption and change as the environment is changing and new origins and destination points or new cities are explored. Moreover if we can explain the emergence of language based on the same sort of cognitive functions required in other domains, then language is less of a mystery as it currently appears.

2. There is increasing evidence from neuroscience that Broca's complex is not exclusively used for language but participates in a wide range of cognitive functions. Broca's complex is perhaps better seen as achieving a more general cognitive function, namely unify structures coming from a mental lexicon triggered by incoming words (Hagoort, 2005). There is also increasing evidence that brain areas that represent perceptual maps or control certain parts of the body become active when language involving these semantic domains is produced or comprehended (Hauk, et.al., 2004). So the notion of a highly specialised modular language organ in the brain is clearly falsified by neuro-scientific evidence and this in turn puts serious doubts about the exclusive role sometimes given to language genes.

3. Models that rely on complex innate structures (like the Principles and Parameters model) or on very language-specific biases (like the Iterated Learning Models) still need to explain where the structures and biases may come from. It seems rather odd for example to suppose that the large inventory of syntactic categories used in languages (the parts of speech, or the syntactic features used in agreement) is innate or strongly biased, particularly because there is increasing doubt that all

these categories are universal across languages Haspelmath(2007). Even in a scenario of genetic assimilation, we must still explain how the behavior patterns and cognitive structures arose that could be assimilated.

Experimental results

The key ingredients of socio-cultural models are: a *population* of individuals, modeled as *agents*, a *micro-world* or ecology which acts as source of meaning for the agents, and a particular type of situated interaction between the agents called a *language game*. An agent has a set of memory structures, for example for storing a lexicon, as well as a set of procedures for carrying out all the information processing that needs to go into language production, comprehension and learning, for example, procedures for parsing a sentence to extract a syntactic structure, or for guessing the meaning of an unknown word and storing it in the lexicon. Besides procedures for producing or interpreting utterances with 'ready-made' linguistic solutions, agents also have diagnostic and repair strategies to handle cases for which they have no solutions yet, and alignment strategies to coordinate their inventories based on the outcome of the language game.

Agents are *autonomous* in the sense that there is no telepathy (one agent cannot inspect or change the state of another agent) and no central control. So there is never any form of direct meaning transfer. There is also no prior 'innate' language or 'innate' conceptual system given by design, because the purpose of the models is precisely to show how these may form and propagate in a population. Agents are *embodied* when the memory and procedures are incorporated in a physical body with sensors and effectors and so interaction takes place within the real world. In *simulated* language games, agents, populations and the world are modeled abstractly and tested in software simulations. In *grounded* language games, the world is the real physical world, the agents are embodied, and the language games involve physical interactions such as pointing gestures or executing actions. Grounded language games can ideally tested through repeatable robotic experiments (see Figure 1).

We have been able to conduct progressively more and more sophisticated robotic experiments of this kind in our laboratory. The sophistication has been growing along several dimensions: the complexity of the robot bodies, the complexity of the micro-world and ecology in which the robots operate, the complexity of the conceptualizations agents are able to construct based on their sensory experiences of the world, the sophistication of their grammars and of the language strategies they use to build and adapt grammars, and consequently the complexity of the emergent language systems

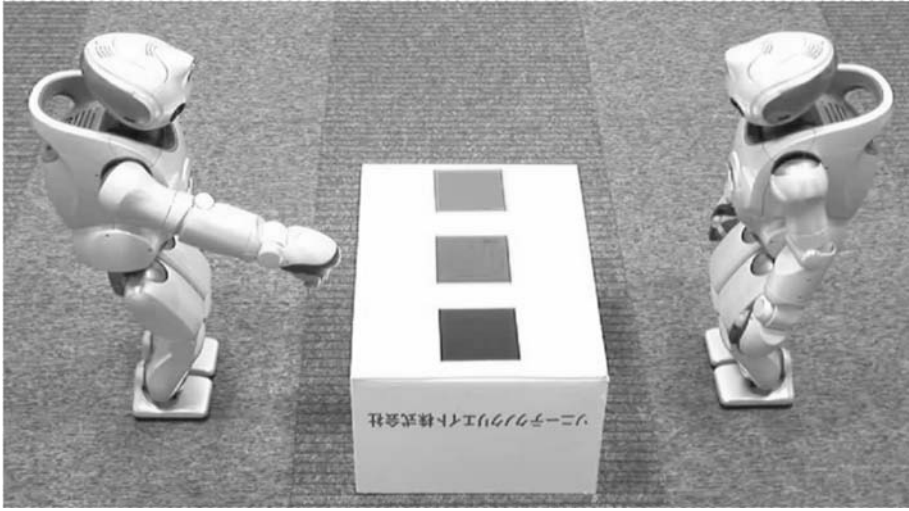


Figure 1. Investigations in the socio-cultural mechanisms underlying language evolution are partially carried out through robotic experiments. Autonomous robots are set up to play language games that require making distinctions (e.g. hue distinctions) and expressing them using words or grammatical constructions. The figure shows a ‘Color Naming Game’ experiment investigating the formation of color categories and color names in a population of humanoid robots

Lexical experiments

The earliest socio-cultural models used simple cybernetic robots (Steels and Vogt, 1997) or pan-tilt cameras (Steels and Kaplan, 2002) acting as ‘Talking Heads’. Research focused mainly on the origins of perceptually grounded categories and lexicons, which is obviously a prerequisite before we can even start to consider grammatical languages. The key questions in these experiments were:

1. Is it possible to pin down the cognitive functions that a group of agents needs in order to settle collectively on a shared set of symbols for naming categories with which they organize their world, such as colors, shapes, or distances? and
2. Is it possible to identify and operationalize the cognitive functions needed to see the emergence of a categorical repertoire in co-evolution with an emergent lexicon? Indeed, if we want to explain the origins of language we need to explain not only the origins of language itself but also the origins of the rich conceptualizations that are expressed by language, and
3. What are the processes that drive the population to a shared lexicon as well as a shared categorical inventory, even though each agent autonomously builds his own.

Investigations into these questions are framed in terms of cooperative interactions called language games (Steels,1995). A language game is a routinized series of turn-taking steps whereby the speaker tries to get the hearer to do something that furthers their joint activities, such as pay attention to a certain object in the world or do a certain action. Playing a language game requires a cooperative attitude and basic skills such as joint attention or agreement on joint scripts. The role of symbols in the language game is comparable to the role of shells or other objects in economic exchange. It does not really matter which symbol is chosen, as long as an agreement is reached on what role it will play in the language game. In current agent-based experiments the scripts are implemented by the experimenter and so far no adequate models have been made how the scripts themselves could arise.

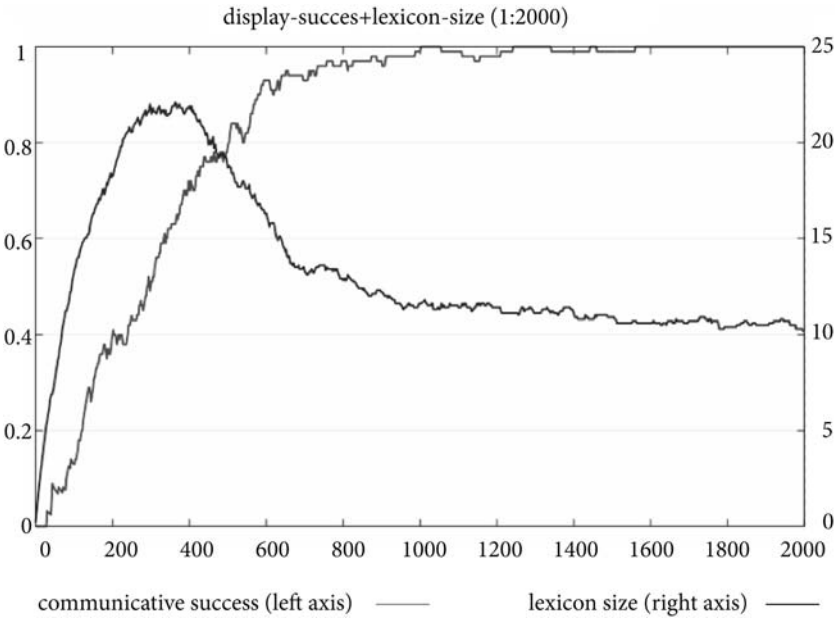


Figure 2. Example of semiotic dynamics generated by language games in a population of 10 agents self-organizing a lexicon to express 10 meanings. The x-axis shows the total number of games played by all agents, each game only involving two agents, and the y-axis communicative success (blue) as well as global size of the inventory (red). The lexicon overshoots initially because agents invent words for certain meanings not knowing that there are already other words for them in the population, but in a second phase agents align their inventories and reach an optimal lexicon of 10 words. We see that agents quickly reach 100% success

It turns out that the cognitive functions that a population of agents requires to self-organise a lexicon are reasonably straightforward and obviously quite generic. They need a bi-directional associative memory with a lateral inhibition dynamics and mechanisms capable of category formation, like radial basis function networks (Steels and Belpaeme, 2005). There must be an invention process to invent new words for uncovered meanings, or for introducing new categories, and a positive feedback loop based on alignment: successful words or categories are enforced. Agents adopt words or categories used by others by guessing possible meanings based on additional non-verbal feedback after a game.

The individual behaviors of each agent create a global effect through self-organization, the same way individual behaviors of ants may lead to global path formation without central control or prior organization. This principle is very familiar to biologists studying collective phenomena, such as the flocking of birds or path formation in ant societies, and physicists that study self-organization in physico-chemical media. Consequently it has been possible to apply the mathematical frameworks developed in complex systems science to the domain of language, showing how and why self-organisation of lexicons works, how such a set-up scales with respect to population size, what the effect is of the social network structure among the agents, how the emergence of a category inventory co-evolves with an emergent lexicon, etc. Loreto and Steels (2007).

In parallel with research on category and lexicon formation, there has been substantial research in my laboratory on the origins of speech systems, see: De Boer (2000), Oudeyer (2006). Similar cognitive functions and interaction patterns as used for the emergence of lexicons and categorical inventories have proven to be effective for evolving an inventory of speech sounds and syllables. These agent-based models for the origins of speech require not only memory structures and categorization and motor control programs, but also a reasonable model of the human articulatory apparatus and of the human hearing system. Instead of a language game, these experiments use an imitation game whereby speaker and hearer try to build a repertoire of 'imitatable sounds'. The agents initially start with random sounds that they then align in order to be more successful in the imitation game. Early research has focused mostly on vowel systems and more research is needed for syllable systems or supra-segmental structures like intonation patterns, but it has now clearly been shown that a sound system can get off the ground fairly easily as soon as agents set themselves to this task and are able to recruit the cognitive functions for it. An emergent sound system can easily be coupled to an emergent lexicon system and an emergent category system, showing a co-evolution of a speech system and a lexical system. Summarising, we can say that the cognitive functions, interaction patterns, and cultural dynamics required for getting a lexical language off the ground and transmitted culturally are reasonably well understood.

Experiments in grammar evolution

Human languages feature not only a vast repertoire of words expressing a huge inventory of perceptually grounded or abstract categories, they add to the lexical items a second layer of information using syntactic means such as word order, morphology, intonation, agreement, hierarchy, or stress patterns, in order to express additional meanings, particularly meanings that go beyond the meaning of individual words such as the role of participants in an event.

More recently, a second wave of socio-cultural, agent-based experiments focusing on grammar has reached a decisive level of performance. In addition to more complex embodiments and more complex visual scene analysis and conceptualization, grammar requires that agents are endowed with capacities to represent lexical and grammatical rules to be used both in parsing and production and that they have the necessary invention, adoption and alignment strategies that lead to a progressive growth and coordination of lexicons and grammars. For our own experiments, we have developed a new linguistic formalism, called Fluid Construction Grammar (or FCG) (Steels(2004), de Beule and Steels(2006)) and shown that it is a sound basis for various experiments in spatial grammar (Steels and Loetzsch(2008)) and case grammar (van Trijp(2008)). We refer for details to this literature and just summarize here some of the important critical features from the viewpoint of the present paper.

Feature structures and inference rules

The representational mechanisms used in FCG are similar to the ones used in problem solvers in general Newell(1990). Language production starts from a particular goal, namely to express as well as possible a particular conceptualisation that will evoke certain behavior in the hearer and language parsing attempts to reconstruct the conceptualisation that fits within the ongoing cooperative activity. The intermediary state in problem solving is represented in terms of feature structures which capture all information about one hypothesis on how the sentence is to be parsed or produced. Lexical entries and grammatical constructions are represented as bi-directional inference rules that can be used to expand a feature structure. Rules are applied using a unification operation that compares the patterns in the conditional part of a rule with the current feature structure and if unification is possible, it will merge the contributing part of the rule to expand the feature structure. Because several rules are usually applicable at every step, parsing and production generates a search space and heuristics are needed to efficiently move through the space. From a mechanistic point of view, linguistic 'problem-solving' is therefore similar to other common sense problem solving such as spatial navigation, action planning and plan recognition, or design, in the sense that

the same computational machinery is adequate even though of course the content of the rules and categories involved is different.

Flexibility and variation

Language users must not only acquire sufficient knowledge of their language so that they can produce or parse sentences, they must also cope with huge variation in the language surrounding them and they must be able to pick out emergent trends in a language and align themselves to these trends. This is achieved in our experiments in various ways. First of all agents store multiple variants of language use in memory, by maintaining scores of the success and frequency of each variant, and by adjusting scores based on success in specific language games using the same lateral inhibition dynamics that was shown to work for the lexicon. Agents should be able to parse or produce sentences, even if certain words are not understood, some grammatical constraints are violated, or adequate ways to express certain aspects of meaning are missing. Because the agent-based models capture the complete cycle of production and comprehension, including perception and action, agents can make up for missing information or constraint violations by falling back on their perceptually derived world models and the task setting.

Meta-level problem solving

Logic-based inference and contemporary computer systems make a distinction between object-level and meta-level processes. The object-level processes carry out routine tasks with ready-made solutions. However when a fail state arises, processing can move to a meta-level in order to try and fix the problem. Such a meta-level architecture has proven to be highly relevant for building effective socio-cultural models of language emergence. Linguistic processing is viewed as a goal-directed problem solving process. The goal of the speaker is to achieve successful communication with minimal cognitive effort, and the goal of the hearer is to cooperate with the hearer to achieve this success. Studies of natural dialog show that natural communication is full of hesitations, false starts, queries for clarification, constant non-verbal feedback so that the speaker can track how well he is understood (e.g. by nodding). There is abundant evidence that speakers self-monitor their ongoing speech and that hearers simulate the language production process of speakers, so that they can complete a sentence or correct a word. The agent-based simulations incorporate these phenomena. The speaker uses himself as a model to simulate the understanding by the hearer and can therefore self-correct or expand his own inventory if necessary. The hearer uses himself as a model of the speaker and can use that model to make reasonable guesses about the meaning and the function of words that he does not know or grammatical constructions that he has

never seen before. By using such meta-level strategies, speakers and hearers constantly expand their linguistic inventories and adapt them to those of others.

Role of cultural evolution

Socio-cultural models do not critically rely on a transmission bottleneck, which is viewed as the primary cause of increased complexity by proponents of iterated learning models Kirby and Hurford(2002). The models are typically tested in two contexts: with and without population change. Experiments without population change show that a language system can arise purely based on the needs of a particular population where members sufficiently interact with each other and functionally improve their language, not only in settings where they share so much of the context that almost no verbal communication is needed but also in settings where the context is only partially shared so that the agents are forced to make more meanings explicit.

Once experiments without population change have been achieved, language game experiments are carried out *with* population change by taking away agents from the population at regular time intervals and by inserting new agents without any knowledge of the existing language. Remarkably, no new mechanisms are needed to explain cultural transmission beyond those explaining how a new language system emerges in a given population in the first place. Generation change will help of course to 'clean up' and further streamline a language system, in the sense that some of the constructs are forgotten that were only rarely used, and of course more and more material is accumulated as successive generations express new conceptual domains or use language in new task settings so that the lexicon and the grammar keep growing. But the socio-cognitive experiments suggest that we do not need to look for a long period of cultural evolution in which the transmission bottleneck operates across generations in order to see the emergence of grammatical language. Children are not considered to be a crucial factor. Both the origins of language and its present-day continued adaptation and expansion is largely the activity of adults, inventing, negotiating and sharing a language system.

Starting a dialog

There is now a significant and growing literature on language game experiments published in specialized journals, particularly in the *Journal of Evolutionary Linguistics*. Contributions can also be found in proceedings of the bi-annual 'Evolution of Language' conferences or in AI and computational linguistics conferences. As research progresses and particularly the tools for doing this Evolutionary Linguistics research become more sophisticated, it becomes more and more feasible to develop evolutionary explanations for realistic language phenomena, such as

Russian aspect or Spanish case grammar, within a reasonable amount of time. These explanations follow a similar methodology as in Evolutionary Biology: A particular language system (like Russian aspect) is deconstructed to find out what language strategy is needed to learn, adapt or invent a similar system, and the communicative function and selective advantage of this strategy is studied through careful systematic experiments.

I now turn to whether we can draw any conclusions from these experiments that may be relevant for understanding how and why humans started to use language. They take the form of social and cognitive prerequisites that had to be built into the language game experiments to get any result at all. These prerequisites are perhaps not as specific as archeologists and anthropologists may want them to be, but they could be a starting point of a dialog between modeling efforts and archeological findings.

Complex categorizations of the world

Language does not express objects or events directly but expresses conceptualizations of the world. The language game experiments all contain an elaborate conceptualization component that is able to come up with meanings that are effective in achieving communicative goals, as well as an interpretation component that takes the meanings reconstructed from sentences and applies them to internal world models. The more sophisticated the conceptualization and interpretation components, the more elaborate phrases can be built.

It follows that an important prerequisite for language is a rich inventory of categorizations and conceptual structuring principles, which of course need not be specific to language although language can help to shape and coordinate categories across members of the population. All animals have a repertoire of perceptually grounded categories that they need for decision-making, recognition of friend and foe, etc. What is unusual for human categorization is the huge amount of categories that we can cope with and the ability to form relatively quickly new categories about any domain relevant to our engagement with the world. So any kind of evidence for the flexible handling of complex categorizations suggests that this particular prerequisite is in place. For example, the handling of complex kinship relations, the ability to deal with a complex changing natural world, navigation skills that require categorization of the weather, landscape patterns, etc.

Symbolism

Language rests fundamentally on the ability of speakers and hearers to deal with symbols. A symbol is an object (sound, gesture, behavior) that has become associated with meaning by convention. The introduction of new words is clearly an act of symbolism. Grammatical constructions are also symbolic. There is no reason

why subjects should come before their verb in a declarative sentence and after them in questions, or why the role of events should be marked with morphological affixes (as in Latin), particles put behind the noun (as in Chinese) or word order and prepositions (as in English). At the core of each language game experiment we always find strategies that language users use to introduce new symbols and coordinate their meanings.

The use of symbols requires the ability to establish interaction patterns and reach agreement about the role of items in such a pattern. Evidence of burials or ritual, evidence of ocre usage, shells acting perhaps as items of economic exchange, are all indications that symbolic behavior has become common in a population. They are not in themselves indicative that language has already emerged but they suggest at least that the symbolism prerequisite is satisfied (d'Errico, F. and L. Backwell (2005)).

Functional thinking for tool use

The language game experiments reported here treat language as a tool. Lexical items and grammatical structures are assumed to have a particular function. When a function is not realized entirely satisfactory, the speaker will extend his language system and he constantly tries to find ways to optimize the amount of effort required in language production and comprehension. The hearer also treats language as a tool, in the sense that if there are unknown words or grammatical markers he will try to find out what their possible meaning and function could be based on problem solving, and then extend his own language system. The speaker is therefore like a tool builder who all the time attempts to improve his tools. The hearer is like somebody who needs to figure out the use of a tool by being able to formulate hypotheses of possible usage based on the context and on success or failures of using them.

Although chimpanzee communities sometimes discover tools (like stones to crack nuts), typical for human tools is the systematic effort to improve them. Tools typically start out in a crude way but then get improved, specialised, manufactured on a larger scale, and culturally shared in a population and across populations. This requires the kind of 'functional thinking' that language game experiments use as well: The tool builder must be able to understand what needs to be achieved, conceive of an initial solution, and observe whether it is performing its desired function. He must make variations to see whether some variant performs better, and then replicate tool construction and propagate tool use to the rest of the community. If language develops as a tool that is constantly being improved and expanded by its users, archeological evidence showing that humans have the inclination and the skill to systematically improve other kinds of tools which leave

clearer traces in the archeological record is an indication that this cognitive prerequisite for language is available (Ambrose(2001)).

Cooperative activities

From a socio-cultural viewpoint, language is deeply embedded in cooperative action. It does not make sense without it. The speaker must want something from the hearer and they both must have an interest to cooperate on some specific task. There must be enough trust between partners in dialog and, definitely in the bootstrapping phase, partners must share enough physical and mental space to be able to make reasonable guesses about the meaning of expressions they have never heard before. These conditions are a strong prerequisite for the language game experiments reported here. The agents are usually programmed to be entirely cooperative. They give honest answers and plan their utterances to maximally help the listener. When a game fails, the speaker will try to be maximally helpful by additional language or non-verbal communication. Although the prerequisite for social cooperation does not need to absolute be within the population as a whole, a language system will collapse or not come of the ground if there is not a minimum number of cooperators or if cheaters cannot be punished as predicted by theories of reciprocal altruism (Wang and Steels, 2008).

Engaging in cooperative action, particularly if it involves guessing what others might want to do, is an extraordinarily complex skill which is of course useful for a broad range of survival tasks, including joint hunting, systematic gathering of food, organisation of living quarters, or defending against danger. Of course many animals have already developed the capacity to band in groups and so these skills have not come out of the blue. Some anthropologists have argued however that humans are unique because they have developed a form of ultra-sociality with respect to non-kin Knight(2007), Boyd and Richerson(2005). And so archeological evidence showing that humans have this kind of strong forms of cooperation can be seen as another strong indicator that the needs and social prerequisites for verbal communication were in place.

Recruitment instead of genetics

I have not said much about neurobiological prerequisites in this paper, because the socio-cultural view of language origins assumes that there is no unique set of specialized language genes nor a highly specialized modular language organ. And so it does not assume a complex innate universal grammar that encodes the categories used in a language system (the parts of speech, the semantic roles, etc.) or the specific patterns that are expected (for example typical ordering of constituents in phrases) because these structures are supposed to emerge based on generic cognitive functions: such as a bi-directional memory, categorization, priming, heuristic

search, pattern matching and instantiation, analogy, perspective reversal, etc. So far the cognitive functions that had to be integrated in computer simulations can easily be imagined to be relevant for a wide variety of other tasks, including spatial navigation or tool design.

If this hypothesis is correct, it suggests that we should *not* be looking for a very specific genetic change that enabled language, a sequence of genetic mutations or recombinations that have shaped a language organ (as suggested by many authors, e.g. Bickerton(1984)). This does not mean there are no biological prerequisites at all. Language needs a broad range of complex cognitive functions and above all a very flexible, reconfigurable brain that can recruit these functions as needed for novel tasks. Language requires a huge amount of processing power and storage capacity. So any biological evolution that pushes the brain in these general directions is a contributing factor. And so evidence of expanding brain size or evidence for increased flexibility so that different human capabilities can be flexibly recruited can be seen as an indication that the neurobiological prerequisites for language are in place.

What all this means for pinpointing the moment when and where humans started on the path to language remains to be clarified by archeological evidence (see d'Errico, et.al. (2003), Davidson(2003)). It appears however that the date at which language may have emerged must be pushed back much further than was believed even a few years ago. Evidence for expanding brain size, flexible recruitment of cognitive functions, extended cooperative group activities, systematic tool improvement, and symbolism has clearly been established for the earliest populations of Homo Sapiens for which evidence is available (Henshilwood, et.al. (2004)). There is very little reason to assume that Neandertals did not have language, and even older hominins show many of the indicators that would correlate with the socio-cultural prerequisites that were necessary in successful language game experiments.

Conclusions

Progressively more complex agent-based models of the emergence and evolution of language are giving us novel insights into the nature of language and language processing and the social and cognitive prerequisites required for language. Early models focused essentially on lexical languages without sophisticated grammar, but recent experiments have shown how increasingly more complex and more human language like structures can emerge, provided the appropriate cognitive and social prerequisites are in place. It now looks like we do not need to put all eggs in the genetic basket, hoping to find a crucial genetic mutation or series of mutations that 'explain' the origins of language. A highly specialised genetically determined language organ is not only increasingly less plausible from a neurobiological point

of view but raises all sorts of difficulties from an evolutionary point of view. In contrast, the socio-cognitive modeling effort shows that given the right social and cognitive mechanisms available for recruitment for language, a communication system with similar properties as human languages can emerge rather quickly in a population of reasonably intelligent individuals and become more and more complex integrating features of human language such as hierarchical structure.

References

- Abrams, D. M. and Strogatz, S. H. (2003) Modelling the dynamics of language death. *Nature*, 424:429.
- Ambrose, S. (2001) Palaeolithic technology and human evolution. *Science* 291, 1748–1753.
- Batali, J. (1998). Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy and C. Knight (Eds.), *Approaches to the Evolution of Language: Social and Cognitive Bases*. Cambridge: Cambridge University Press.
- Bickerton, D. (1984) The language bioprogram hypothesis. *Behavioral and Brain Sciences* 7, 17–388.
- Boyd, R. and P. Richerson (2005) *The Origins and Evolution of Culture*. Oxford: Oxford University Press.
- Cangelosi A., D. Parisi (1998). The emergence of a “language” in an evolving population of neural networks. *Connection Science*, 10(2), 83–9.
- Davidson, I. (2003) The archeological evidence of language origins: State of the Art. In: M. Christiansen and S. Kirby (eds.) *Language Evolution*, pp. 140–157. Oxford: Oxford University Press.
- De Beule, J. and L. Steels (2005) Hierarchy in Fluid Construction Grammar. In: Furbach, U. (ed.) *Proceedings of KI-2005*, pp 1–15. Berlin: Springer.
- De Boer, B. (2000) Self organization in vowel systems. *Journal of Phonetics* 28(4): 441–465.
- d’Errico, F. and L. Backwell (eds.) (2005) *From Tools to Symbols. From Early Hominids to Modern Humans*. Witwatersrand University Press.
- d’Errico, F., C. Henshilwood, G. Lawson, M. Vanhaeren, A.-M. Tillier, M. Soressi, F. Bresson, B. Maureille, A. Nowell, J. Lakarra, C. Backwell, and M. Julien, (2003). Archaeological evidence for the emergence of language, symbolism and music. An alternative multidisciplinary perspective. *Journal of World Prehistory* 17: 1–70.
- Elman, J. L., Bates, E.A., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
- Griffiths, T. and Kalish, M. (2005). A Bayesian view of language evolution by iterated learning. In B. G. Bara, L. Barsalou and M. Bucciarelli (Eds.), *Proceedings of the XXVII Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hagoort, P. (2005) On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*. 9(9) 416–423.
- Haspelmath, M. (2007) Pre-established categories don’t exist: consequences for language description and typology. *Linguistic Typology* 11.1:119–132
- Hauk, O., Johnsrude, I., and Pulvermuller, F. (2004). Somato-topic representation of action words in human motor and pre-motor cortex. *Neuron*, 41, 301–307.

- Heine, B. (1997) Cognitive foundations of grammar. Oxford: Oxford University Press.
- Heine, B. and T. Kuteva (2008) *The Genesis of Grammar. A reconstruction*. Oxford: Oxford University Press
- Henshilwoord, C., F. d'Errico, R. Yates, Z. Jacobs, C. Tribolo, N. Duller, J. Sealy, H. Valladas, I. Watts and A. Wintle (2002) Emergence of modern human behaviour: Middle Stone Age engravings from South Africa. *Science* (295): 1278–1280.
- Holland, J.H. (1992) *Adaptation in Natural and Artificial Systems*. Cambridge, MA: MIT Press.
- Hopper, P. (1987) Emergent grammar. *Berkeley Linguistics Conference (BLS)*, 13:139–157.
- Kirby, S. and Hurford, J. (2002) The Emergence of Linguistic Structure: An overview of the Iterated Learning Model. In Angelo Cangelosi and Domenico Parisi, editors, *Simulating the Evolution of Language*, pp 121–148. London: Springer.
- Knight, C. (2007) Language co-evolved with the rule of law. *Mind and Society*. 7(1), 109–128.
- Loreto, V. and L. Steels (2007) Emergence of language. *Nature Physics* 3(11), 758–760.
- Lyon, C. and Nehaniv, C. L. and Cangelosi, A. (ed.) (2007) *Emergence of Language and Communication*, pp 129–151. Berlin: Springer.
- Minett, J. W. and W. S-Y. Wang (2005) *Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics*. City University of Hong Kong Press: Hong Kong.
- Newell, A. (1990) *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.
- Oudeyer, P-Y (2006) *Self-organization in the evolution of speech*. Oxford: Oxford University Press.
- Pinker, S. and R. Jackendoff (2005) The Faculty of Language: What's Special about it? *Cognition* 95, 201–236
- Steels, L. (1995) A Self-organizing Spatial Vocabulary. *Artificial Life* 2(3), pp. 319–332.
- Steels, L. (2003) Evolving grounded communication for robots. *Trends in Cognitive Science*. 7(7), 308–312.
- Steels, L. (2004) Constructivist Development of Grounded Construction Grammars Scott, D., Daelemans, W. and Walker M. (eds) *Proceedings Annual Meeting Association for Computational Linguistic Conference*. Barcelona. p. 9–19.
- Steels, L. (2005) The Emergence and Evolution of Linguistic Structure: From Lexical to Grammatical Communication Systems. *Connection Science*. 17(3).
- Steels, L. (2007) The Recruitment Theory of Language Origins. In: Lyon, C. and Nehaniv, C. L. and Cangelosi, A. (ed.) *Emergence of Language and Communication*, pp 129–151. Berlin: Springer.
- Steels, L. and T. Belpaeme (2005) Coordinating Perceptually Grounded Categories through Language. A Case Study for Colour. *Behavioral and Brain Sciences*. 24(6), ??-??.
- Steels, L., F. Kaplan, A McIntyre and J. Van Looveren (2002) Crucial factors in the origins of word-meaning. In Wray, A. (ed.) *The Transition to Language*. Oxford: Oxford University Press.
- Steels, L. and M. Loetzsch (2008) Perspective Alignment in Spatial Language. In: Coventry, K., J. Bateman and T. Tenbrink (eds) *Spatial Language in Dialogue*. Oxford: Oxford University Press.
- Steels, L. and Vogt, P. (1997) Grounding Adaptive Language Games in Robotic Agents. In Harvey, I. and P. Husbands, P. (eds.) *Proceedings of the 4th European Conference on Artificial Life*, pp. 474–482. Cambridge, MA: The MIT Press.
- Szathmari, E. (2001) Origin of the human language faculty: the language amoeba hypothesis. In: J. Trabant and S. Ward (eds) *New Essays on the Origins of Language*. (Trends in Linguistics, Studies and Monographs, 133). pp. 41–55. Berlin: Mouton de Gruyter.

- Szathmary, E. (2007) Towards and understanding of language origins. In: M. Barbieri (ed.) *Codes of Life*, pp. 283–313. Berlin: Springer.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
- Thornton, R. and K. Wexler (1999) *Principle B, VP Ellipsis, and Interpretation in Child Grammar*. Cambridge, MA: The MIT Press.
- van Trijp, Remi (2008). The Emergence of Semantic Roles in Fluid Construction Grammar. In: Smith, Andrew D.M., Kenny Smith and Ramon Ferrer i Cancho (eds.) *The Evolution of Language. Proceedings of the 7th International Conference*. pp. 346–353. Singapore: World Scientific Publishing.
- Wang, E. and L. Steels (2008) Self-interested agents can bootstrap symbolic communication if they punish cheaters. In: Smith, Andrew D.M., Kenny Smith and Ramon Ferrer i Cancho (eds.) *The Evolution of Language. Proceedings of the 7th International Conference*. Singapore: World Scientific Publishing.

Index

A

Afrolittorina africana 34, 35
 agent-based models 268, 275, 277, 282
 aggregate models 268
 AGR 104–109, 111
 agriculturalist 112, 115, 173
 alarm calls 240–243, 245, 247–249, 251, 252, 254–256
 Altaic 163, 164, 166, 182
 Altamura 36
 Alu sequences 123, 127, 129, 140
 ambient language 205–207, 210, 211, 213, 216, 219–221, 225–228
 Amud 36, 37, 45
 ancient DNA 75, 76, 81
 ape 257
 archaeology 11, 13, 15, 46, 69, 71, 74, 75, 93, 100, 224
 Arcy-sur-Cure 25, 51
 articulation 45, 208, 210, 213, 215–218, 226, 236, 239, 240
 Assam 147–150, 153, 155–157, 160, 161, 183
 associative memory 275
 Aurochs 69, 71, 74, 80–83
 Austroasiatic 182
 Autosomal markers 127, 129, 166, 171, 173, 176
 Autosomal Short Tandem Repeat (STR) loci 182

B

Bantu 93–115
 Bantu expansion 93, 95, 107–109, 111, 114, 115
 Bantu-speaking 93–95, 100, 103, 104, 106–114
 Beçov 21, 22
 Berekhat Ram 29
 Bhutan 181–187, 190–192, 194
 Bilzingsleben 28
 Black Mountain Mönpa 185, 187, 188

Blombos Cave 22, 24, 27–30, 32, 33, 49
 Bodo 38, 161, 185, 186,
 body size 72, 81, 239
 Boro-Garo languages 153–156, 158, 159
 Brassempouy 38
 Broca 45, 271
 Budongo 245, 256
 burial 13, 35–39
 Burushaski 182

C

Cameroon 93–95, 100, 102, 104, 111, 112, 114
 canonical babbling 206–208, 211
 Carpenter's Gap 30
 categories 19, 149, 216, 220, 223, 224, 237, 250, 270–273, 275–277, 279, 281
 cattle 69–71, 73–75, 78, 80–84
 Cavallo 25
 Central Asia 76, 163–166, 173, 177–178
 Chatelperronian 25, 27, 35, 48
 chimpanzees 19, 44, 235, 237, 240, 254, 256–258
 China 182, 186, 189, 192
 cognitive capacities 44, 207, 247, 268, 270
 cognitive functions 267, 270, 271, 273, 275, 281, 282
 complexity 16, 41, 47, 48, 206, 207, 209, 225–227, 272, 278
 Congo-Brazzaville 94
 consonant 160, 208–210, 213–215, 217, 222, 231
 contact 47–49, 69, 83, 94, 96, 101, 102, 115, 116, 161, 163, 164, 237, 238
 convergence 95–98, 100–102
 cooperation 69, 184, 193, 194, 281
 cooperative action 281
 cooperative interactions 274

cosmetic coalition model 19
 Crayon 20
 Cro Magnon 38
 cross language 218, 220–227

D

Daic 182
 Democratic Republic of Congo 104, 105
 demography 50, 107, 139, 148, 182
 Denisova 35
 depictions 13, 14, 30
 descent systems 116
 dialect continuum 152, 156
 Diana monkeys 240, 249–252, 254–256
 diffusion 65, 67, 69–71, 73, 76–78, 80, 85
 Divje Babe 42
 domestication 70–76, 80–83, 85, 138
 Dravidian 182, 191

E

eagle 240, 241, 243, 245–247, 249, 251, 252
 early words 212, 214, 215, 217, 222
 Eastern Asia 165, 181, 183, 192
 El Sidron 25
 encephalisation 14
 Ermitage 29
 Europe 14, 15, 21, 23, 25, 28, 30, 34–36, 38, 39, 41, 47, 69–71, 76–85, 130, 131, 171
 evolutionary linguistics 278

F

Fang 95, 96, 102, 103, 111, 113, 115
 farming 69–71, 76, 78, 84, 85, 94–96, 98, 102, 112–114
 Fluid Construction Grammar 276, 283
 food 31, 78, 115, 116, 148, 155, 235, 257, 281

- FOXP2 25, 50, 242
 frame-content 208, 221, 226
 front-back 213, 220–222
 FST 106, 107, 134, 171, 187, 190, 191
- G**
 Gabon 93–96, 98, 99, 101–104, 107, 111, 113, 114, 116
 Geissenklösterle 39–42
 genetic and linguistic diversity 163, 164, 166
 genetic markers 108, 113, 123, 127, 136, 190
 genetics 16, 69, 71, 74, 83, 117, 118, 121, 164, 177, 181–182, 190, 195, 196, 224, 268, 281
 gibbon 44
 Glycymeris insubrica 34
 goat 69–72, 74–80, 84
 grave goods 14, 36, 37
 Gravettian 38–41
 Grotte des Pigeons 30–32, 34
 Grotte du Renne 25, 48
 grunts 235, 257
 guinea fowls 254
- H**
 haplogroup 79, 83, 104–109, 111, 114, 115, 131–138
 Haa Pteah 42
 height 213, 220–222, 226
 Herto 38
 Himalayan range 147
 Himalayan region 181–183, 190, 192
 Hohle Fels 39–41
 Homo erectus 44, 46, 240
 Homo ergaster 46
 H. sapiens 15, 23, 25, 29, 30, 34, 36–39, 43, 44, 47, 48
 human migration 79, 181, 183
 Human Revolution 15, 47
 hunter-gatherers 50, 69, 94–96, 98, 99, 104, 106–111, 113, 114
- I**
 India 147, 150, 154, 155, 161, 182–184, 186, 187, 189, 190, 192
 Indo-European 163, 164, 166, 171, 176, 182, 185, 187, 188, 190–192
 Indo-Iranian 163, 165–167, 171–173, 175–177
 inductive learning 269
 innate structures 271
 input 39, 165, 187, 206, 207, 210, 211, 216, 223, 226–228
 intersyllabic 208, 209, 226
 intonation 210, 275, 276
 Isturitz 38–42
- K**
 Kara-Bom 35
 Kebara 36, 37, 43
 Khotyk 35
 kinship 155, 172, 238, 270, 279
 Klasies 24, 28
 Klein Kliphuis 28
 Knapping 16, 73, 77
 Kostienki 38
 Kriging 187
 Kusunda 182
- L**
 La Chapelle-aux-Saints 29, 37, 43
 La Combe 38
 La Ferrassie 29, 37
 language 11, 13–21, 25, 30, 36, 42–48, 93, 95, 96, 98–103, 113–115, 123–127, 138–140, 150, 152, 153, 155–158, 160, 163, 164, 166, 167, 170–173, 175, 181–187, 190–194, 205–214, 216, 218–233, 235, 236, 242, 245, 247, 256, 258–259, 267–283
 language evolution 206, 224, 236, 247, 267, 269, 273, 283
 language game 272, 274, 275, 278–282
 language history 158
 language replacement 101, 113
 language shift 99, 113, 125, 155, 176
 larynx 43–45, 237, 239–241, 258
 later words 212, 214, 215
 learning 206, 207, 209, 210, 216, 220, 225–227, 237, 249, 259, 269–272, 278, 283
 Le Moustier 37
 leopard 240, 241, 243, 249, 251, 252, 254–257
 Le Piage 39
 Les Pradelles 29
 Levenstein distance 163, 175
 lexicon 46, 103, 114, 116, 158, 205, 211, 267, 271–275, 277, 278
 Lhokpu 185, 187, 188, 194
 lifestyle 94, 115, 116, 176
 linguistic borders 181
 linguistic distance 163, 164, 170, 190
- M**
 Malakunanja II 27
 Maloyalomanskaia 35
 Mandu Mandu 35
 Manganese 20, 23–26, 28
 manner of articulation 208, 213, 215, 217, 218, 226
 Marillac 29
 mating patterns 94, 114, 116
 Melacortin 25
 memory 46, 238, 270, 272, 275, 277, 281
 migrations 69, 95, 124, 127, 137, 139, 143
 MIS 4 49, 50
 MIS 5 49, 50
 mitochondrial DNA 82, 104, 123, 127, 130, 131, 166, 172, 173
 Mladeč 39
 modeling 224, 225, 267–270, 279, 283
 monkeys 235, 237, 238, 240, 242–252, 254–256, 258
 MtDNA 78, 80, 83, 93, 95, 103–106, 108, 109, 111, 113, 114, 130–132, 137, 169, 175, 183, 190, 192
 multilingualism 101, 113
 musical instruments 14, 42
 musilanguage 17
- N**
 Naga 149–152, 161
 natural selection 269, 271
 Nauwalabila 27
 navigation 270, 271, 276, 279, 282
 Nazlet Kather 38
 Neanderthals 20, 23–25, 27, 30, 34, 36, 38, 39, 42, 43, 45–48, 50
 Neolithic 39, 70, 71, 73–85, 102, 124, 137, 165
 Neolithic dissemination 70
 Nepal 181–185, 187, 190–192
- O**
 Oase 39
 ochre 20, 22, 23, 25, 27–29, 33, 38
 ontogeny 206, 207, 210, 225, 227–228

- ornaments 13, 14, 20, 30, 34, 35,
38, 47–49
Oued Djebbana 30, 32, 34
Out-of-Africa 15, 16, 47, 48
- P**
pant hoot 237
paternal lineage 108, 109, 115
patrilocality 94, 113, 173
Pech de l'Azé 23, 24
Pekary Ila 27
perception 103, 209, 210, 236,
270, 277
PHG 104–109, 111
phylogeny 44, 133, 206, 207,
209, 227
pigment 13, 19–23, 25, 30, 31, 47
Pinnacle Point 21, 24, 67
place of articulation 213, 217,
218, 226
Podzvonkaya 35
population density 94, 95,
148–150
population movements 93, 94,
116, 137, 153, 182
Porosphaera globularis 34
predator 240–245, 247, 249–251,
254, 257
primates 43, 46, 235–242, 244,
245, 247–249, 254, 257, 258
problem solving 269–271, 276,
277, 280
production 13, 22, 45, 47, 49,
78, 152, 155, 205–210, 212, 219,
220, 224–228, 236–238, 242,
244, 247, 258, 270, 272, 276,
277, 280
Pygmies 93, 94, 98, 100,
105–109, 114
Pygmy 93, 94, 96, 104, 106–111,
114, 237
pyows 242, 243, 245, 246, 251
- Q**
Qafzeh 23, 24, 29, 30, 33, 34,
36–38
Qena 38
Quneitra 29
- R**
Ras el Kelb 34
reduplication 210
Riwi 35
robot 272
- robotic experiments 267, 270,
272, 273
- S**
Sai Island 21, 22
Saint-Césaire 38
Sefunim 34
sheep 4, 69–72, 74–78, 80, 84
Sibudu 30, 34, 35, 47–49
signals 8, 230, 235, 237, 248, 249
Sima de los Huesos 38, 43
simulations 9, 267–270, 272,
277, 282
Skhul 30, 32, 34, 36
social memory 238
Sodmein 23, 24
squirrels 237, 249
structure 106, 110, 128, 131,
137, 158, 160, 173, 186–190,
210, 236, 237, 239–241, 257,
268–272, 275, 276, 283
supra-segmental structures 275
Swadesh list 170
syllable systems 275
symbolic 4, 15, 17–21, 27–30, 35,
36, 46–48, 50, 85, 206, 268,
279, 280
symbolism 15, 18, 20, 35, 46, 85,
279, 280, 282
syntactic structures 270
- T**
Tabun 37
Taforalt 30, 31, 49
Tāi 240, 243, 247, 254
Tani 149–152, 157
Tan-Tan 30
technologies 14, 41, 50, 176
technology 3, 15, 17, 42, 116
Temnata 28
Tibet 6, 147, 192
Tibeto-Burman 181, 182, 185–192
Toto 185–188, 194, 200
Turkic 163, 165–167, 171–173,
175–177
Twin Rivers 21, 47
- U**
Uçagizli 34
Uluzzian 25, 48
Unikote 29
- V**
variegated babbling 220
Vaufrey 29
- Vergisson 29
vocal tract 8, 14, 43, 45, 52, 218,
237, 239–241
Vogelherd 39, 41
vowels 7, 96, 158, 160, 208, 209,
213–215, 219, 220, 226
vowel systems 275
- W**
Wernicke 45
- Y**
Y-chromosome 5, 6, 93, 95, 103,
108–111, 113–115, 123, 127, 133,
135–137, 166, 169, 172, 173,
175, 190